



**AFRL-RY-WP-TR-2010-1075**

**ANALYSIS OF PARAMETRIC ADAPTIVE SIGNAL  
DETECTION WITH APPLICATIONS TO RADARS AND  
HYPER SPECTRAL IMAGING**

**Hongbin Li**

**Stevens Institute of Technology  
Department of Electrical and Computer Engineering**

**FEBRUARY 2010  
Final Report**

**Approved for public release; distribution unlimited.**

*See additional restrictions described on inside pages*

**STINFO COPY**

**AIR FORCE RESEARCH LABORATORY  
SENSORS DIRECTORATE  
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320  
AIR FORCE MATERIEL COMMAND  
UNITED STATES AIR FORCE**

## NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the Wright-Patterson Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RY-WP-TR-2010-1075 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

\*//Signature//

BRADLEY J. MORTENSON, Work Unit Manager  
Electronics Engineer, AFRL/RVRT

//Signature//

WILLIAM J. BALDYGO  
Branch Chief, AFRL/RVRT

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

\*Disseminated copies will show “//Signature//” stamped or typed above the signature blocks.

<b>REPORT DOCUMENTATION PAGE</b>					<i>Form Approved</i> OMB No. 0704-0188				
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>									
<b>1. REPORT DATE (DD-MM-YY)</b> February 2010		<b>2. REPORT TYPE</b> Final		<b>3. DATES COVERED (From - To)</b> 16 December 2004 – 15 February 2010					
<b>4. TITLE AND SUBTITLE</b> ANALYSIS OF PARAMETRIC ADAPTIVE SIGNAL DETECTION WITH APPLICATIONS TO RADARS AND HYPERSPECTRAL IMAGING				<b>5a. CONTRACT NUMBER</b>					
				<b>5b. GRANT NUMBER</b> FA8750-05-2-0001					
				<b>5c. PROGRAM ELEMENT NUMBER</b> 62204F					
<b>6. AUTHOR(S)</b> Hongbin Li				<b>5d. PROJECT NUMBER</b> 5017					
				<b>5e. TASK NUMBER</b> RL					
				<b>5f. WORK UNIT NUMBER</b> 517R1507					
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Stevens Institute of Technology Department of Electrical and Computer Engineering Castle Point on Hudson Hoboken, NJ 07030-5591				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>					
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Air Force Research Laboratory Sensors Directorate Wright-Patterson Air Force Base, OH 45433-7320 Air Force Materiel Command United States Air Force				<b>10. SPONSORING/MONITORING AGENCY ACRONYM(S)</b> AFRL/RVRT					
				<b>11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)</b> AFRL-RY-WP-TR-2010-1075					
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> Approved for public release; distribution unlimited.									
<b>13. SUPPLEMENTARY NOTES</b> PAO Case Number: 88ABW 2010-1162; Clearance Date: 15 March 2010. This report contains color.									
<b>14. ABSTRACT</b> New parametric space-time adaptive processing (STAP) based detectors are introduced and examined. Unlike conventional techniques that estimate the characteristics of the disturbance from only the secondary data, the proposed detectors obtain such knowledge jointly from the primary and secondary data. When the number of pulses within a coherent processing interval is sufficiently large, the proposed detectors can function even without any secondary data, making them strong candidates for detection in non-homogeneous environments. The proposed detectors are investigated by both theoretical analysis and numerical study using simulated and real radar data. Extensive comparison with conventional STAP methods shows that the proposed detectors can better deal with training-limited scenarios while being computationally simpler. The proposed techniques are also extended for target detection in hyperspectral imaging (HSI).									
<b>15. SUBJECT TERMS</b> space-time adaptive processing, parametric modeling, radar target detection, non-homogeneous environment, hyperspectral imaging									
<b>16. SECURITY CLASSIFICATION OF:</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;"><b>a. REPORT</b> Unclassified</td> <td style="padding: 2px;"><b>b. ABSTRACT</b> Unclassified</td> <td style="padding: 2px;"><b>c. THIS PAGE</b> Unclassified</td> </tr> </table>			<b>a. REPORT</b> Unclassified	<b>b. ABSTRACT</b> Unclassified	<b>c. THIS PAGE</b> Unclassified	<b>17. LIMITATION OF ABSTRACT:</b> SAR		<b>18. NUMBER OF PAGES</b> 142	
<b>a. REPORT</b> Unclassified	<b>b. ABSTRACT</b> Unclassified	<b>c. THIS PAGE</b> Unclassified							
<b>19a. NAME OF RESPONSIBLE PERSON (Monitor)</b> Bradley J. Mortenson			<b>19b. TELEPHONE NUMBER (Include Area Code)</b> N/A						

# Table of Contents

<b>1 Introduction .....</b>	<b>1</b>
1.1 Background.....	1
1.2 Summary of Work.....	3
<b>2 Review of Space-time Adaptive Processing .....</b>	<b>6</b>
2.1 Phased Array Radar.....	6
2.2 Non-parametric STAP Approaches.....	7
2.2.1 Matched Filter Detector.....	7
2.2.2 Sample Matrix Inversion Approach.....	8
2.2.3 Adaptive Matched Filter Detector.....	8
2.2.4 Kelly's GLRT.....	8
2.2.5 Adaptive Coherence Estimator Detector.....	9
2.2.6 Drawbacks of Non-parametric STAP Detectors.....	9
2.3 Parametric STAP Approaches.....	9
2.3.1 Parametric Adaptive Matched Filter Detector.....	9
2.3.2 Swindlehurst and Stoica's GLRT.....	10
<b>3 Parametric Rao Test for Multichannel Adaptive Signal Detection.....</b>	<b>11</b>
3.1 Introduction.....	11
3.2 Data Model and Problem Statement.....	12
3.3 The Parametric Rao Test.....	14
3.3.1 Test Statistic.....	14
3.3.2 Asymptotic Analysis.....	16
3.4 Numerical Results.....	17
3.5 Conclusions.....	19
<b>4 Parametric GLRT for Multichannel Adaptive Signal Detection.....</b>	<b>25</b>
4.1 Introduction.....	25
4.2 Data Model and Problem Statement.....	26
4.3 Parameter Estimation.....	28
4.3.1 ML Estimation under $H_1$ .....	28
4.3.2 ML Estimation under $H_0$ .....	29
4.3.3 Asymptotic ML Estimation under $H_1$ .....	30
4.3.4 CRB.....	31
4.4 Parametric GLRT.....	32
4.4.1 Test Statistic.....	32
4.4.2 Asymptotic Analysis.....	33
4.5 Numerical Results.....	33
4.5.1 Estimation.....	34
4.5.2 Detection.....	34
4.6 Conclusions.....	35
<b>5 A Simplified Parametric GLRT for Multichannel Adaptive Signal Detection .....</b>	<b>40</b>
5.1 Introduction.....	40
5.2 Data Model and Parametric GLRT.....	41
5.2.1 Data Model.....	41

## Table of Contents, continued

5.2.2 Parametric GLRT.....	42
5.3 Amplitude Estimation.....	44
5.4 New Parametric GLRT.....	46
5.5 Numerical Examples.....	47
5.5.1 Estimation.....	47
5.5.2 Detection.....	48
5.5.3 KASSPER Dataset.....	48
5.6 Conclusion.....	49
<b>6 Recursive Parametric Tests for Multichannel Adaptive Signal Detection .....</b>	<b>53</b>
6.1 Introduction.....	53
6.2 Data Model and Problem Statement.....	54
6.3 Non-recursive Parametric Rao and Parametric GLRT Detectors with Known $P$ .....	55
6.4 Recursive Parametric Tests with Unknown $P$ .....	58
6.4.1 Parameter Estimation by the Multichannel Levinson Algorithm.....	58
6.4.2 AR Model Order Selection.....	60
6.4.3 Recursive Parametric Rao Test.....	60
6.4.4 Recursive Parametric GLRT.....	61
6.4.5 Complexity.....	62
6.5 Numerical Results.....	62
6.5.1 Estimation.....	63
6.5.2 Detection.....	63
6.6 Conclusions.....	64
<b>7 Performance Evaluation of Parametric Space-Time Adaptive Detectors .....</b>	<b>68</b>
7.1 Introduction.....	68
7.2 Data Model and Problem Statement.....	69
7.3 Parametric Rao and GLRT Detectors.....	70
7.4 Performance Evaluation.....	73
7.4.1 Simulated KASSPER Dataset.....	73
7.4.2 Measured MCARM Dataset.....	75
7.4.3 Bistatic Dataset.....	77
7.5 Conclusions.....	79
<b>8 Parametric Adaptive Signal Detection for Hyperspectral Imaging .....</b>	<b>88</b>
8.1 Introduction.....	88
8.2 Data Model and Problem Statement.....	89
8.3 Covariance-Matrix Based Solutions.....	91
8.4 Proposed Approach.....	92
8.4.1 Parametric Modeling of HIS Data.....	92
8.4.2 NS-AR Model Based Parametric Adaptive Detectors.....	95
8.4.3 ML Estimation of NS-AR Coefficients.....	96
8.4.4 NS-AR Model Order Selection.....	98
8.4.5 Training Screening.....	99
8.5 Experimental Results.....	100
8.5.1 Model Order Selection.....	101

## Table of Contents, concluded

8.5.2 Detection in Homogeneous Environments.....	101
8.5.3 Detection in Heterogeneous Environments.....	102
8.6 Conclusions.....	102
<b>A Parametric Rao Test .....</b>	<b>109</b>
A.1 ML Parameter Estimation.....	109
A.2 Derivation of the Parametric Rao Test.....	111
A.3 Asymptotic Distribution of the Parametric Rao Test Statistic.....	114
A.4 Performance of the MF and AMF Detectors.....	114
<b>B Parametric GLRT .....</b>	<b>116</b>
B.1 ML Parameter Estimation.....	116
B.2 Derivation of the AML Estimator.....	118
B.3 Unbiasedness and Consistency of the LS Estimator.....	118
B.4 Derivation of CRB.....	119
B.5 Asymptotic Distribution of the Parametric GLRT Statistic.....	120
<b>CA Simplified Parametric GLRT .....</b>	<b>121</b>
C.1 Derivation of (5.23).....	121
C.2 Derivation of the Amplitude Estimator.....	122
C.3 Derivation of the New Parametric GLRT.....	123
C.4 Alternative Form of the Parametric GLRT.....	124

# Chapter 1

## Introduction

### 1.1 Background

Multichannel signal detection is encountered in a wide variety of applications. In radar systems, sensor arrays are often used to facilitate the so-called *space-time adaptive processing* (STAP), which offers enhanced target discrimination capability compared with space- or time-only processing [1, 2]. In remote sensing systems, multispectral and hyperspectral sensors are used to collect spectral information across multiple spectral bands, which can be exploited for classification of different materials or detection of man-made objects on the ground [3, 4]. Other examples of applications include wireless communications, geolocation, sonars, audio and speech processing, and seismology [5–7].

STAP based multichannel signal detectors have been successfully utilized to mitigate the effects of clutter and/or interference in radar, remote sensing, and communication systems [1–5]. The first work on STAP for radar was by Brennan and Reed [8] where optimum space-time filtering was presented. In [9], the Reed, Mallett, and Brennan (RMB) detector was proposed. They substitute the sample covariance matrix for the true covariance matrix in the optimal solution. This RMB detector is implemented as follows: 1) estimate a covariance matrix from target-free training data, 2) determine a weight vector from the obtained covariance matrix estimate, 3) form a test statistic and compare it with a threshold for signal detection. This RMB detector is not a constant false alarm rate (CFAR) detector, since the probability of false alarm depends on the true covariance matrix. Simple modification to this RMB detector leads to the CFAR adaptive matched filter (AMF) detector [10–12]. In addition, there are several well-known STAP detectors such as Kelly’s generalized likelihood ratio test (GLRT) [13], the adaptive coherence estimator (ACE) detector [14–16], and so forth.

However, the aforementioned conventional STAP detectors involve estimating and inverting a large-size space-time covariance matrix of the disturbance signal (viz., clutter, jamming, and noise) for each test cell using target-free training data. This may impose excessive training and computational burdens when the joint space-time dimension is large. At a minimum, we need

$K \geq JN$  training data<sup>1</sup> to ensure a full-rank estimate of the  $JN \times JN$  space-time covariance matrix, where  $J$  denotes the number of spatial channels and  $N$  the number of temporal observations. Moreover, the RMB rule [9] suggests that, in average,  $K \geq (2JN - 3)$  training data are needed to obtain performance within 3 dB from the optimum bound. Such conditions may not be satisfied, especially in heterogeneous (due to varying terrain, high platform altitude, bistatic geometry, conformal array, among others) or dense-target environments that offer limited training, thus rendering covariance matrix based techniques inapplicable.

The problem of target detection with *limited training* data support has gained great attention in recent decades [17–21]. Typical cases in airborne radar systems include dense-target and/or heterogeneous environments. In either case, the detection performance of the conventional STAP detectors degrades significantly because of the mismatch of the space-time covariance matrix estimate relative to that of the target test cell.

Parametric model based STAP detectors have also gained significant interest in recent years. [22–25]. In many applications, however, the disturbance usually exhibits certain spatial, temporal, and/or spectral structures that can be exploited to reduce the number of unknowns and ease the training/computational burden. Among other alternatives, one general structured approach is to model the disturbance as a multichannel autoregressive (AR) process, which has been found to be very useful in representing the spatial and temporal correlation of radar signals [22–25]. A parametric detector based on such a multichannel AR model was developed in [22, 23], which is referred to as the *parametric adaptive matched filter* (PAMF) detector. The PAMF detector has been shown to significantly outperform the aforementioned covariance matrix-based detectors for small training size at reduced complexity. Specifically, the PAMF detector models the disturbance as a multichannel AR process driven by a temporally white but spatially colored multichannel noise. While traditional STAP detectors perform joint space-time whitening (using an estimate of the space-time covariance matrix), the PAMF detector adopts a two-step approach that involves temporal whitening via an inverse moving-average (MA) filter followed by spatial whitening. The parameters that need to be estimated are the AR coefficient matrices and the spatial covariance matrix of the driving multichannel noise, which are significantly fewer than what are involved in estimating the space-time covariance matrix. This is the essence behind the training and computational efficiency of the PAMF detector.

Although intuitively sound, the PAMF detector was obtained in a heuristic approach by modifying the AMF test statistic. Specifically, it replaces the joint space-time whitening incurred by the AMF detector with two separate whitening procedures in time and space as discussed above. The test threshold, false alarm, and detection probabilities were determined primarily by computer simulation, due to the limited analysis of the PAMF detector. Moreover, the aforementioned detectors use only the target-free training data for the estimation of the space-time covariance matrix. Then, a test statistic is formed and compared with a test threshold. Thus, it renders the detectors inapplicable in the case where the target-free training data are unavailable.

---

<sup>1</sup> $K \geq JN - 1$  training data are needed if both the test and training data are used to estimate the space-time covariance matrix.



## 1.2 Summary of Work

This report examines the problem of detecting a multichannel signal in the presence of spatially and temporally colored disturbance. The main research objective is to develop multichannel parametric detectors by modeling the disturbance as a multichannel AR process and exploiting this parametric model in a GLRT approach. The joint probability density functions (PDFs) of the test and training data under the null and alternative hypotheses are used to develop the parametric detectors, just like what Kelly did in his seminal paper [13].<sup>2</sup>

A first contribution of this report is the development of a *parametric Rao test* for the multichannel signal detection problem. A generic Rao test is known to offer a standard solution to a class of parameter testing problems. It is easier to derive and implement than the GLRT, and is also asymptotically (large-sample in the number of temporal observations and/or training data) equivalent to the latter. The Rao test was recently used to develop detectors for several other problems [26, 27]. A detailed discussion on the attributes of a generic Rao test can be found in [7].

Our parametric Rao test differs from the generic one for multichannel signal detection in that we make explicit use of a multichannel AR model for the disturbance signal. We show that, interestingly, the parametric Rao test takes a form identical to that of the PAMF detector. The only difference is that we use a maximum likelihood (ML) based estimator that involves using both test and training signals for parameter estimation, whereas the estimators in [23] use only training signals for that purpose. If the ML estimator is utilized, the parametric Rao/PAMF detector is asymptotically a parametric GLRT. The asymptotic distribution of the test statistic under both hypotheses is obtained in closed form, which can be used to set the test threshold and compute the corresponding detection and false alarm probabilities. Since the asymptotic distribution under the null hypothesis is independent of the unknown parameters, the parametric Rao/PAMF detector asymptotically achieves constant false alarm rate (CFAR). Numerical results are presented, which show that our asymptotic results are accurate in predicting the performance of the Rao/PAMF detector even when the data size is modest.

A second contribution of this report is the development of a *parametric GLRT*. It is natural to extend the results of the parametric Rao test and consider the parametric GLRT for several reasons. First, the problem of interest is a two-sided parameter testing problem that admits no uniformly most powerful (UMP) solution [7]. A GLRT approach is widely used in such cases due to its good asymptotic properties including asymptotic CFAR and consistency. Second, the parametric GLRT may yield improved performance than the parametric Rao detector, especially when the data is limited. This is because the latter is an asymptotic (large-sample) parametric GLRT [7, Appendix 6B]. Third, all Rao tests, including the parametric Rao detector, are obtained based on a further approximation that is valid only for weak signals [7, p. 238]. As such, the parametric Rao detector is expected to degrade when the weak signal assumption is violated. The above observations motivate us to consider the parametric GLRT, in hope of finding a better solution to the problem.

By developing the parametric Rao and GLRT detectors, we also investigate the underlying

---

<sup>2</sup>It is noted that Kelly's GLRT does not utilize the parametric model for the disturbance.

ML parameter estimation. The parametric GLRT relies on ML parameter estimation for both the null and alternative hypotheses while the parametric Rao test depends only on ML parameter estimation under the null hypothesis. We show that the ML estimator under the alternative hypothesis is non-linear and requires searches on a two-dimensional parameter space. To address this issue, we introduce an asymptotic ML (AML) estimator that is considerably simpler, yielding estimates in a non-iterative fashion, and asymptotically coincides with the optimum ML estimator. The Cramér-Rao Bound (CRB) for the estimation problem is also derived, offering a baseline for comparing various (unbiased) estimators.

A third contribution of this report is a *simplified parametric GLRT*. The parametric GLRT detector involves a highly nonlinear maximum likelihood estimation procedure, which was solved via a two-dimensional iterative search method initialized by a suboptimal estimator [28]. To facilitate the detection, a simplified GLRT along with a new estimator is proposed. Both the estimator and the GLRT are derived in closed-form at considerably lower complexity. With adequate training data, the new GLRT achieves a similar detection performance as the original one. However, for the more interesting case of limited training, the original GLRT may become inferior due to poor initialization. Because of its simpler form, the new GLRT also offers additional insight into the parametric multichannel signal detection problem. The performance of the proposed detector is assessed using both a simulated dataset, which was generated using multichannel AR models, and the KASSPER dataset, a widely used dataset with challenging heterogeneous effects found in real-world environments.

A fourth contribution of this report is the *recursive implementation* of parametric Rao and parametric GLRT detectors. The parametric Rao and parametric GLRT detectors assume that the model order of the multichannel AR process is known a priori to the detector. In practice, the model order has to be estimated by some model order selection technique. Meanwhile, a standard non-recursive implementation of the parametric detectors is computationally intensive since the unknown parameters have to be estimated for all possible model orders before the best one is identified. To address these issues, we consider joint model order selection, parameter estimation, and target detection problem. We present recursive versions of the aforementioned parametric detectors by integrating the multichannel Levinson algorithm, which is employed for recursive and computationally efficient parameter estimation [29,30], with a generalized Akaike Information Criterion (GAIC) for model order selection [30,31]. Numerical results show that the proposed recursive parametric detectors yield a detection performance nearly identical to that of their non-recursive counterparts at significantly reduced complexity.

A fifth contribution of this report is performance evaluation of parametric space-time adaptive detectors with more challenging datasets. The parametric detectors were studied under the assumption that the disturbance follows a multichannel AR process. However, the disturbance signals in an airborne radar environment do not necessarily follow an exact multichannel AR model. In this chapter, the parametric Rao and GLRT detectors are studied using a number of datasets: 1) simulated data - Knowledge-Aided Sensor Signal Processing and Expert Reasoning (KASSPER) dataset that was generated by a high-fidelity clutter model; 2) measured data - Multi-Channel Airborne Radar Measurement (MCARM) dataset that was collected from an L-band airborne phased array radar testbed; and 3) bistatic data that was generated based on a

bistatic geometry. These datasets contain many real-world effects such as heterogeneous terrains, antenna errors and leakage, dense ground targets/discretes, and range-dependent clutter (bistatic geometry). Experimental results show that the parametric detectors can provide good detection performance with limited or no range training even in more realistic radar environments.

The last chapter of this report talks about application of parametric adaptive signal detection to *hyperspectral imaging*. Traditional approaches to the so-called subpixel target signal detection problem are *training-inefficient* due to the need for an estimate of a large-size covariance matrix of the background from target-free training pixels. This imposes a training requirement that is often difficult to meet in a heterogeneous environment. A class of training-efficient adaptive signal detectors is presented by exploiting a parametric model that takes into account the non-stationarity of HSI data in the spectral dimension. A maximum likelihood (ML) estimator is developed to estimate the parameters associated with the proposed parametric model. Several important issues are discussed, including model order selection, training screening, and time-series based whitening and detection, which are intrinsic parts of the proposed parametric adaptive detectors. Experimental results using real HSI data reveal that the proposed parametric detectors are more training-efficient and outperform conventional covariance-matrix based detectors when the training size is limited.

The remainder of this report is organized as follows. Chapter 2 briefly reviews the STAP for airborne radar detection. Chapters 3, 4 and 5 present the parametric Rao test, parametric GLRT and the simplified parametric GLRT, respectively. Details of the technical developments of the results reported in Chapters 3, 4 and 5 are found in Appendices A, B and C, respectively. Chapter 6 presents the recursive versions of the parametric detectors. Performance evaluation of parametric space-time adaptive detection with more challenging datasets is introduced in Chapter 7. And Chapter 8 discusses application of parametric adaptive signal detection to hyperspectral imaging.

*Notation:* Vectors (matrices) are denoted by boldface lower (upper) case letters, all vectors are column vectors, superscripts  $(\cdot)^T$ ,  $(\cdot)^*$ , and  $(\cdot)^H$  denote transpose, complex conjugate, and complex conjugate transpose, respectively,  $\mathbf{I}_M$  denotes the  $M \times M$  identity matrix (with subscript suppressed sometimes),  $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$  denotes the multivariate complex Gaussian distribution with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\mathbf{R}$ ,  $\|\cdot\|$  takes the Frobenius norm of a matrix/vector,  $\otimes$  denotes the Kronecker product,  $\text{vec}(\cdot)$  denotes the operation of stacking the columns of a matrix on top of each other,  $\mathbb{C}$  denotes the complex number field,  $\Re\{\cdot\}$  takes the real part of the argument, and  $\Im\{\cdot\}$  takes the imaginary part, and finally,  $(\cdot)^\dagger$  denotes the Moore-Penrose pseudo-inverse.

# Chapter 2

## Review of Space-time Adaptive Processing

Adaptive temporal processing has been used for cancellation of weather clutter [32]. Adaptive spatial processing or array processing has been used for interference nulling in radar. Adaptive array processing techniques can provide nulling far below the sidelobe level limitation, but special measures must be taken to avoid the inclusion of mainlobe clutter in an adaptive beamforming radar [1].

STAP simultaneously combines and processes signals received on multiple channels (the spatial domain) and from multiple pulse repetition periods (the temporal domain). Therefore, it is useful whenever the received signals (target and interference) are functions of both space and time [32]. STAP is an efficient tool for slow target detection in strong clutter environments, and provides robustness to system errors and capability to handle non-stationary interference [1].

In this chapter, we focus on STAP for airborne radar detection. However, it should be noted that STAP has broad applications in radar, sonar, remote sensing, and communication systems [1–5]. In addition, the research results obtained in this dissertation can also be applicable to other areas, e.g., sonar, biomedical imaging, remote sensing, wireless communications, and so on.

### 2.1 Phased Array Radar

A linear phased array airborne radar with  $J$  antennas is considered herein [1]. The radar transmits a coherent burst of  $N$  pulses at a constant pulse repetition frequency (PRF)  $f_r = 1/T_r$ , where  $T_r$  is the pulse repetition interval (PRI). The transmitter frequency is  $f_o = c/\lambda_o$ , where  $c$  is the propagation velocity and  $\lambda_o$  is the radar operating wavelength. The waveform returns are collected over the time interval, referred to as the coherent processing interval (CPI). For each PRI,  $K_T$  time (range) samples are collected to cover the range interval. The received data is organized in a  $J \times N \times K_T$  CPI data cube.

The test data, or primary data consist of  $J \times N$  samples from a single range gate. These samples are arranged in a column vector  $\mathbf{x}_0$ . The training data, or secondary data consist of  $K$  range gates, forming a subset of the  $K_T - 1$  remaining ones. These are denoted by  $\mathbf{x}_k$ ,  $k = 1, 2, \dots, K$ .

Given the test and training data, the detection problem can be expressed as a binary hypothesis testing:

$$\begin{aligned} H_0 : \quad \mathbf{x}_0 &= \mathbf{d}_0, \\ H_1 : \quad \mathbf{x}_0 &= \alpha \mathbf{s} + \mathbf{d}_0, \end{aligned} \quad (2.1)$$

where  $\mathbf{s}$  is the target steering vector,  $\alpha$  is an *unknown* amplitude, and  $\mathbf{d}_0$  is the *disturbance data* that may be correlated in space and time. In addition to the test data  $\mathbf{x}_0$ , there may be a set of *training data*  $\mathbf{x}_k$ ,  $k = 1, \dots, K$ , that are target-free:  $\mathbf{x}_k = \mathbf{d}_k$ . The disturbance data  $\{\mathbf{d}_k\}_{k=0}^K$  are assumed to be independent and identically distributed (i.i.d.) with distribution  $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ , where  $\mathbf{R}$  is the *unknown* space-time covariance matrix.

For a uniform equi-distant linear array, the steering vector takes the form [23]:

$$\mathbf{s} = \mathbf{s}_s(\omega_s) \otimes \mathbf{s}_t(\omega_d), \quad (2.2)$$

where  $\mathbf{s}_s(\omega_s)$  denotes the  $J \times 1$  spatial steering vector:

$$\mathbf{s}_s(\omega_s) = \frac{1}{\sqrt{J}} [1, e^{j\omega_s}, \dots, e^{j\omega_s(J-1)}]^T, \quad (2.3)$$

and  $\mathbf{s}_t(\omega_d)$  denotes the  $N \times 1$  temporal steering vector:

$$\mathbf{s}_t(\omega_d) = \frac{1}{\sqrt{N}} [1, e^{j\omega_d}, \dots, e^{j\omega_d(N-1)}]^T, \quad (2.4)$$

where  $\omega_s$  and  $\omega_d$  denote the normalized target spatial and Doppler frequencies, respectively.

## 2.2 Non-parametric STAP Approaches

The space-time processor linearly combines all the samples from the range gate of interest and produces a scalar output [1]:

$$T = |\mathbf{w}^H \mathbf{x}_0|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \quad (2.5)$$

where  $\gamma$  is the test threshold.

The space-time processor consists of three major steps. The first step is the training strategy which selects the training data from the given data cube. The second step is the weight vector computation. Based on the training data, the adaptive weight vector is computed. The final step is the weight vector application, which computes a scalar output, or test statistic. The output of the processor is compared with a test threshold to determine whether a target is present or not.

### 2.2.1 Matched Filter Detector

If the space-time covariance matrix, denoted by  $\mathbf{R}$  is known exactly, the optimum detector that maximizes the output SINR is the matched filter (MF) [12]:

$$T_{\text{MF}} = |\mathbf{w}_{\text{MF}}^H \mathbf{x}_0|^2 = \frac{|\mathbf{s}^H \mathbf{R}^{-1} \mathbf{x}_0|^2}{\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{MF}}, \quad (2.6)$$

where

$$\mathbf{w}_{\text{MF}} = \frac{\mathbf{R}^{-1}\mathbf{s}}{\sqrt{\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}}}, \quad (2.7)$$

$\gamma_{\text{MF}}$  denotes the MF threshold. The MF detector is obtained by a GLRT approach (e.g., [7]), by which the ML estimate of the unknown amplitude  $\alpha$  is first estimated and then substituted back into the likelihood ratio to form a test statistic. It should be noted that the MF detector *cannot* be implemented in real applications since  $\mathbf{R}$  is unknown. However, it provides a baseline for performance comparison when considering any realizable detection scheme.

### 2.2.2 Sample Matrix Inversion Approach

In practice, the unknown  $\mathbf{R}$  should be replaced by some estimate, such as the sample covariance matrix obtained from the secondary data [9]:

$$\hat{\mathbf{R}} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H. \quad (2.8)$$

Using  $\hat{\mathbf{R}}$  and the adaptive weight vector obtained by

$$\mathbf{w}_{\text{RMB}} = \hat{\mathbf{R}}^{-1} \mathbf{s}, \quad (2.9)$$

gives the RMB detector [9]. It is also known as the sample matrix inversion (SMI) approach. The RMB detector, however, lacks CFAR.

### 2.2.3 Adaptive Matched Filter Detector

A modification to the RMB detector leads to the so-called AMF detector [10–12]:

$$T_{\text{AMF}} = \frac{|\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{x}_0|^2}{\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{AMF}}, \quad (2.10)$$

where  $\gamma_{\text{AMF}}$  denotes the AMF threshold. It is noted that the AMF test statistic exhibits a CFAR property.

### 2.2.4 Kelly's GLRT

Alternatively, one can treat both  $\alpha$  and  $\mathbf{R}$  as unknowns and estimate them successively by ML. Such a GLRT approach was pursued by Kelly [13], which gives the following Kelly test:

$$T_{\text{Kelly}} = \frac{|\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{x}_0|^2}{\left(\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}\right) \left(K + \mathbf{x}_0^H \hat{\mathbf{R}}^{-1} \mathbf{x}_0\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{Kelly}}, \quad (2.11)$$

where  $\gamma_{\text{Kelly}}$  denotes the corresponding threshold.

### 2.2.5 Adaptive Coherence Estimator Detector

Conte et al. [33, 34] and Scharf et al. [14] have proposed another GLRT based detection scheme which is the so-called Adaptive Coherence Estimator (ACE) detector:

$$T_{\text{ACE}} = \frac{\left| \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{x}_0 \right|^2}{\left( \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} \right) \left( \mathbf{x}_0^H \hat{\mathbf{R}}^{-1} \mathbf{x}_0 \right)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{ACE}}, \quad (2.12)$$

where  $\gamma_{\text{ACE}}$  denotes the ACE threshold. Clearly, the ACE detector is a normalized version of the AMF detector. It is noted that the ACE test statistic lies between 0 and 1.

### 2.2.6 Drawbacks of Non-parametric STAP Detectors

The AMF, ACE, and Kelly' GLRT detectors have a CFAR property, which is a highly desirable feature of STAP detectors. However, they also entail a large training requirement. In particular, the sample covariance matrix  $\hat{\mathbf{R}}$  has to be inverted, which imposes a constraint on the training size

$$K \geq JN \quad (2.13)$$

to ensure a full-rank  $\hat{\mathbf{R}}$ . The RMB rule [9] suggests that at least

$$K \geq (2JN - 3) \quad (2.14)$$

target-free secondary data vectors are needed to obtain an expected performance within 3 dB from the optimum MF detector. Such a training requirement may be difficult to meet, especially in non-homogeneous or dense-target environments. Besides excessive training, the computational complexity of these detectors is also high, since  $\hat{\mathbf{R}}$  has to be computed and inverted for each CPI.

## 2.3 Parametric STAP Approaches

The aforementioned difficulties are primarily because for large  $JN$  and in the absence of any specific structure,  $\mathbf{R}$  involves an enormous number of unknowns. In many applications, however, the disturbance usually exhibits certain spatial, temporal, and/or spectral structures that can be exploited to reduce the number of unknowns and ease the training/computational burden. Among other alternatives, one general structured approach is to model the disturbance as a multichannel AR process, which has been found to be very useful in representing the spatial and temporal correlation of radar signals [22–25].

### 2.3.1 Parametric Adaptive Matched Filter Detector

Using both simulated and real data, the PAMF detector [22, 23] has been shown to significantly outperform the aforementioned covariance matrix-based detectors for small training data size at

a reduced complexity. While covariance matrix-based detectors perform joint space-time whitening for interference mitigation, the PAMF detector adopts a two-step approach, involving temporal whitening via an inverse moving-average (MA) filter followed by spatial whitening.

The PAMF detector is given by [23]

$$T_{\text{PAMF}} = \frac{\left| \sum_{n=P}^{N-1} \hat{\mathbf{s}}_P^H(n) \hat{\mathbf{Q}}_P^{-1} \hat{\mathbf{x}}_{0,P}(n) \right|^2}{\sum_{n=P}^{N-1} \hat{\mathbf{s}}_P^H(n) \hat{\mathbf{Q}}_P^{-1} \hat{\mathbf{s}}_P(n)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{PAMF}}, \quad (2.15)$$

where  $\hat{\mathbf{Q}}_P$  denotes an estimate of the spatial covariance matrix  $\mathbf{Q}$ ,  $\hat{\mathbf{x}}_{0,P}(n)$  and  $\hat{\mathbf{s}}_P(n)$  are the *temporally* whitened test signal and steering vector, respectively; which are whitened using an inverse AR filter of order  $P$  (i.e., a multichannel MA filter) whose parameters, along with  $\hat{\mathbf{Q}}_P$ , are estimated from the secondary data. The parametric approach offers savings in both training and computation, since the parameters to be estimated are significantly fewer, compared with covariance matrix-based approaches.

However, when the training data are subject to outlier contamination, the performance of the PAMF detector degrades severely. Moreover, the PAMF detector was obtained in a heuristic approach by modifying the AMF test statistic. Specifically, it replaces the joint space-time whitening incurred by the AMF detector with two separate whitening procedures in time and space as discussed earlier. The test threshold, false alarm, and detection probabilities were determined primarily by computer simulation, due to the limited theoretical analysis available for the PAMF detector.

### 2.3.2 Swindlehurst and Stoica's GLRT

Swindlehurst et al. have presented the GLRT in [24]. They considered a different detection problem that involves unknown non-linear signal parameters associated with the signal to be detected. In their problem, the steering vector is parameterized by unknown Doppler and direction of arrival (DOA). They have also utilized the vector AR model for the disturbance, which is spatially and temporally correlated. Their vector AR (VAR) approach uses only training data for parameter estimation. In other words, they treat estimation and detection as separate processes, as opposed to the joint processing used in Kelly's GLRT.



# Chapter 3

## Parametric Rao Test for Multichannel Adaptive Signal Detection

### 3.1 Introduction

In this chapter, we develop a *parametric Rao test* for the multichannel signal detection. A generic Rao test is known to offer a standard solution to a class of parameter testing problems. It is easier to derive and implement than the GLRT, and is also asymptotically (large-sample in the number of temporal observations and/or training data) equivalent to the latter. The Rao test was recently used to develop detectors for several other problems [26, 27]. A detailed discussion on the attributes of a generic Rao test can be found in [7].

Our parametric Rao test differs from the generic one for multichannel signal detection in that we make explicit use of a multichannel AR model for the disturbance signal. We show that, interestingly, the parametric Rao test takes a form identical to that of the PAMF detector. The only difference is that we use a maximum likelihood (ML) based estimator that involves using both test and training signals for parameter estimation, whereas the estimators in [23] use only training signals for that purpose. If the ML estimator is utilized, the parametric Rao/PAMF detector is asymptotically a parametric GLRT. Under the conditions stated in Section 7.2, the asymptotic distribution of the test statistic under both hypotheses is obtained in closed form, which can be used to set the test threshold and compute the corresponding detection and false alarm probabilities. Since the asymptotic distribution under  $H_0$  is independent of the unknown parameters, the parametric Rao/PAMF detector asymptotically achieves constant false alarm rate (CFAR). Numerical results are presented, which show that our asymptotic results are accurate in predicting the performance of the Rao/PAMF detector even when the data size is modest.

The remainder of this chapter is organized as follows. Section 7.2 contains the data model and problem statement. Our main results are summarized in Section 3.3. In particular, Section 3.3.1 contains a summary of the parametric test statistic, while Section 3.3.2 includes our asymptotic analysis. Details of the technical developments of the results reported in Section 3.3 are found in Appendices A.1 to A.3. Numerical results are presented in Section 3.4 followed by concluding

remarks in Section 3.5.

## 3.2 Data Model and Problem Statement

The problem under consideration involves detecting a known multichannel signal with unknown amplitude in the presence of spatially and temporally correlated disturbance (e.g., [1]):

$$\begin{aligned} H_0 : \quad & \mathbf{x}_0(n) = \mathbf{d}(n), \quad n = 0, 1, \dots, N-1, \\ H_1 : \quad & \mathbf{x}_0(n) = \alpha \mathbf{s}(n) + \mathbf{d}(n), \quad n = 0, 1, \dots, N-1, \end{aligned} \quad (3.1)$$

where all vectors are  $J \times 1$  with  $J$  denoting the number of spatial channels, and  $N$  is the number of temporal observations. Henceforth,  $\mathbf{x}_0(n)$  is called the *test signal*,  $\mathbf{s}(n)$  is the signal to be detected with amplitude  $\alpha$ , and  $\mathbf{d}(n)$  is the *disturbance signal* that may be correlated in space and time. In addition to the test signal, it is assumed that a set of target-free *training* or *secondary* data vectors  $\mathbf{x}_k(n)$ ,  $k = 1, 2, \dots, K$  and  $n = 0, 1, \dots, N-1$ , are available to assist in the signal detection process.

Define the following  $JN \times 1$  space-time vectors:

$$\begin{aligned} \mathbf{s} &= [\mathbf{s}^T(0), \mathbf{s}^T(1), \dots, \mathbf{s}^T(N-1)]^T, \\ \mathbf{d} &= [\mathbf{d}^T(0), \mathbf{d}^T(1), \dots, \mathbf{d}^T(N-1)]^T, \\ \mathbf{x}_k &= [\mathbf{x}_k^T(0), \mathbf{x}_k^T(1), \dots, \mathbf{x}_k^T(N-1)]^T, \\ &\quad k = 0, 1, \dots, K. \end{aligned} \quad (3.2)$$

Equation (7.1) can be more compactly written as

$$\begin{aligned} H_0 : \quad & \mathbf{x}_0 = \mathbf{d}, \\ H_1 : \quad & \mathbf{x}_0 = \alpha \mathbf{s} + \mathbf{d}. \end{aligned} \quad (3.3)$$

Clearly, the composite hypothesis testing problem (7.1) or (3.3) is also a *two-sided parameter testing problem* that tests  $\alpha = 0$  against  $\alpha \neq 0$ . The general assumptions in the literature are (e.g., [1, 9–13, 15, 16, 22, 23]),

- **AS1:** The signal vector  $\mathbf{s}$  is deterministic and *known* to the detector;
- **AS2:** The signal amplitude  $\alpha$  is complex-valued, deterministic, and *unknown*;
- **AS3:** The secondary data  $\{\mathbf{x}_k\}_{k=1}^K$  and the disturbance signal  $\mathbf{d}$  (equivalently,  $\mathbf{x}_0$  under  $H_0$ ) are independent and identically distributed (i.i.d.) with distribution  $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ , where  $\mathbf{R}$  is the *unknown* space-time covariance matrix.

In particular, the above signal detection problem occurs in an airborne STAP radar system with  $J$  array channels and a coherent processing interval (CPI) of  $N$  pulse repetition intervals (PRIs). The disturbance  $\mathbf{d}(n)$  consists of ground clutter, jamming, and thermal noise, while

$\mathbf{s}(n)$  is called the *target space-time steering vector*. For a uniform equi-spaced linear array,<sup>1</sup> the steering vector is given by [23]:

$$\mathbf{s}(\omega_s, \omega_d) = \mathbf{s}_s(\omega_s) \otimes \mathbf{s}_t(\omega_d), \quad (3.4)$$

where  $\mathbf{s}_s(\omega_s)$  denotes the  $J \times 1$  spatial steering vector:

$$\mathbf{s}_s(\omega_s) = \frac{1}{\sqrt{J}} [1, e^{j\omega_s}, \dots, e^{j\omega_s(J-1)}]^T \quad (3.5)$$

and  $\mathbf{s}_t(\omega_d)$  denotes the  $N \times 1$  temporal steering vector:

$$\mathbf{s}_t(\omega_d) = \frac{1}{\sqrt{N}} [1, e^{j\omega_d}, \dots, e^{j\omega_d(N-1)}]^T, \quad (3.6)$$

where  $\omega_s$  and  $\omega_d$  denote the normalized target spatial and Doppler frequencies, respectively.

While Assumptions **AS1** to **AS3** are standard (e.g., [1, 9–13, 15, 16]), we further assume the following:

- **AS4**: The disturbance signal  $\mathbf{d}(n)$  can be modeled as a multichannel  $\text{AR}(P)$  process with *known* model order  $P$  but *unknown* AR coefficient matrices and spatial covariance (see *Remark 1* below for additional comments on this assumption).

Based on Assumption **AS4**, the secondary data  $\{\mathbf{x}_k\}_{k=1}^K$  are represented as

$$\begin{aligned} \mathbf{x}_k(n) &= - \sum_{p=1}^P \mathbf{A}^H(p) \mathbf{x}_k(n-p) + \boldsymbol{\varepsilon}_k(n), \\ k &= 1, 2, \dots, K, \end{aligned} \quad (3.7)$$

where  $\{\mathbf{A}^H(p)\}_{p=1}^P$  denote the  $J \times J$  AR coefficient matrices, and  $\boldsymbol{\varepsilon}_k(n)$  denote the driving multi-channel spatial noise vectors that are temporally white but spatially colored Gaussian noise:  $\boldsymbol{\varepsilon}_k(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ , where  $\mathbf{Q}$  denotes the  $J \times J$  spatial covariance matrix. Meanwhile, the test signal  $\mathbf{x}_0$  is given by

$$\begin{aligned} \mathbf{x}_0(n) &= \alpha \mathbf{s}(n) \\ &= - \sum_{p=1}^P \mathbf{A}^H(p) \{\mathbf{x}_0(n-p) - \alpha \mathbf{s}(n-p)\} + \boldsymbol{\varepsilon}_0(n), \end{aligned} \quad (3.8)$$

where  $\alpha = 0$  under  $H_0$ ,  $\alpha \neq 0$  under  $H_1$ , and  $\boldsymbol{\varepsilon}_0(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ . Let  $\tilde{\mathbf{s}}(n)$  denote a regression on  $\mathbf{s}(n)$  and  $\tilde{\mathbf{x}}_0(n)$  a regression on  $\mathbf{x}_0(n)$  under  $H_1$ :

$$\tilde{\mathbf{s}}(n) = \mathbf{s}(n) + \sum_{p=1}^P \mathbf{A}^H(p) \mathbf{s}(n-p), \quad (3.9)$$

$$\tilde{\mathbf{x}}_0(n) = \mathbf{x}_0(n) + \sum_{p=1}^P \mathbf{A}^H(p) \mathbf{x}_0(n-p). \quad (3.10)$$

---

<sup>1</sup>Note that the results presented in this section apply to any array configurations, as long as the steering vector is known (cf. Assumption **AS1**).

Then, the driving noise in (3.8) can be alternatively expressed as

$$\varepsilon_0(n) = \tilde{\mathbf{x}}_0(n) - \alpha\tilde{\mathbf{s}}(n). \quad (3.11)$$

The problem of interest is to develop a decision rule for the above composite hypothesis testing problem using the test and training signals as well as exploiting the multichannel parametric AR model.

*Remark 1:* We shall clarify that our goal here is not to justify whether AR models are appropriate or not for STAP applications. An answer to the question can be found in [23], where it is shown that low-order multichannel AR models are very powerful and efficient in capturing the temporal and spatial correlation of the disturbance and, hence, can greatly help signal detection in airborne STAP systems. As stated above, our problem is how to exploit such a parametric model to solve the composite testing problem. The assumption that the model order  $P$  is known is only used to simplify our presentation. In practice, the model order has to be estimated, and a variety of model order selection techniques, such as the Akaike Information Criterion (AIC) and the Minimum Description Length (MDL) based techniques (e.g., [35] and references therein), are available for this task. Since such techniques may over- or under-estimate the true model order, a relevant problem is how the proposed detector performs when over- or under-estimation occurs (also see [36]). This will be investigated in Section 3.4. Finally, it is also possible to formulate the problem to include  $P$  as another parameter to be estimated. We do not follow such an approach in order to focus on the relations between the parametric Rao test and the PAMF detector, which also assumes that a prior estimate of  $P$  is available.

## 3.3 The Parametric Rao Test

### 3.3.1 Test Statistic

The derivation of the parametric Rao test that takes into account Assumptions **AS1** to **AS4** in Section 7.2 is presented in Appendix A.2, which in turns relies on the ML estimates of the nuisance parameters (i.e., parameters associated with the disturbance signal) that are obtained in Appendix A.1. The test is given by<sup>2</sup>

$$T_{\text{Rao}} = \frac{2 \left| \sum_{n=P}^{N-1} \hat{\mathbf{s}}^H(n) \hat{\mathbf{Q}}^{-1} \hat{\mathbf{x}}_0(n) \right|^2}{\sum_{n=P}^{N-1} \hat{\mathbf{s}}^H(n) \hat{\mathbf{Q}}^{-1} \hat{\mathbf{s}}(n)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{Rao}}, \quad (3.12)$$

where  $\gamma_{\text{Rao}}$  denotes the test threshold, which can be set by using the results in Section 3.3.2,  $\hat{\mathbf{s}}(n)$  and  $\hat{\mathbf{x}}_0(n)$  denote, respectively, the steering vector and test signal that have been whitened temporally, and additional spatial whitening is provided by  $\hat{\mathbf{Q}}^{-1}$ , which is the inverse of the ML estimate of the spatial covariance matrix to be specified next.

---

<sup>2</sup>Although the factor of 2 on the test statistic can be absorbed by the test threshold, it is retained to keep the asymptotic distribution of the test statistic more compact. See Section 3.3.2.

Specifically, the temporally whitened steering vector and test signal in (7.14) are obtained as follows:

$$\hat{\mathbf{s}}(n) = \mathbf{s}(n) + \sum_{p=1}^P \hat{\mathbf{A}}^H(p) \mathbf{s}(n-p), \quad (3.13)$$

$$\hat{\mathbf{x}}_0(n) = \mathbf{x}_0(n) + \sum_{p=1}^P \hat{\mathbf{A}}^H(p) \mathbf{x}_0(n-p), \quad (3.14)$$

where  $\hat{\mathbf{A}}^H(p)$  denotes the ML estimate of the AR coefficient matrix  $\mathbf{A}^H(p)$ .

To present the ML estimates more compactly, let

$$\mathbf{A}^H = [\mathbf{A}^H(1), \mathbf{A}^H(2), \dots, \mathbf{A}^H(P)] \in \mathbb{C}^{J \times JP}, \quad (3.15)$$

which contains all the coefficient matrices involved in the  $P$ -th order AR model, and

$$\mathbf{y}_k(n) = [\mathbf{x}_k^T(n-1), \mathbf{x}_k^T(n-2), \dots, \mathbf{x}_k^T(n-P)]^T, \quad (3.16)$$

$$k = 0, 1, \dots, K,$$

which contains the regression subvectors formed from the test signal  $\mathbf{x}_0$  or the  $k$ -th training signal  $\mathbf{x}_k$ . We first compute the following correlation matrices:

$$\hat{\mathbf{R}}_{xx} = \sum_{n=P}^{N-1} \sum_{k=0}^K \mathbf{x}_k(n) \mathbf{x}_k^H(n), \quad (3.17)$$

$$\hat{\mathbf{R}}_{yy} = \sum_{n=P}^{N-1} \sum_{k=0}^K \mathbf{y}_k(n) \mathbf{y}_k^H(n), \quad (3.18)$$

$$\hat{\mathbf{R}}_{yx} = \sum_{n=P}^{N-1} \sum_{k=0}^K \mathbf{y}_k(n) \mathbf{x}_k^H(n). \quad (3.19)$$

Then, the ML estimates of the AR coefficients  $\mathbf{A}^H$  and the spatial covariance matrix  $\mathbf{Q}$  are given by (see Appendix A.1)

$$\hat{\mathbf{A}}^H = -\hat{\mathbf{R}}_{yx}^H \hat{\mathbf{R}}_{yy}^{-1}, \quad (3.20)$$

$$\hat{\mathbf{Q}} = \frac{1}{(K+1)(N-P)} \left( \hat{\mathbf{R}}_{xx} - \hat{\mathbf{R}}_{yx}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{R}}_{yx} \right). \quad (3.21)$$

*Remark 2:* The PAMF detector also involves estimating the AR coefficients  $\mathbf{A}^H$  and the spatial covariance matrix  $\mathbf{Q}$  [23]. Several estimators were suggested, including the Strand-Nuttall algorithm and the least-squares (LS) estimators. The LS estimator was observed to yield better performance. Our ML estimator is similar to the LS estimator except that we use both the test and training signals to obtain parameter estimates, whereas the latter utilizes only the training signals for parameter estimation. A subscript ‘‘P’’ is therefore used for the parameter estimates in

(2.15) to indicate the difference. Note that with the ML estimator, it is possible to derive parameter estimates exclusively from the test signal, thus obviating the need for training. This could be advantageous especially in highly heterogeneous environments where it is difficult to obtain training signals that are i.i.d. with respect to the disturbance in the test signal. The detection performance of the parametric Rao detector in the absence of training will be explored elsewhere. We would like to point out that our approach is similar to Kelly's GLRT [13], which also employs both the test and training signals for parameter estimation. However, we shall stress that Kelly's GLRT does not exploit the multichannel parametric model as shown in (7.3) and (3.8).

*Remark 3:* By comparing the parametric Rao test statistic (7.14) and the PAMF test statistic (2.15), we can quickly see that if both detectors use the ML estimator for parameter estimation, they are identical except for a scaling factor of 2. Hence, under the conditions stated in Section 7.2, *the PAMF detector is a parametric Rao detector*. Since the parametric Rao test is asymptotically equivalent to the *parametric GLRT*<sup>3</sup>, the PAMF detector, with the ML parameter estimates, is also an asymptotic parametric GLRT. As we shall see in Section 3.3.2, the equivalence offers additional insights into the performance and implementation of the PAMF detector.

*Remark 4:* It should be noted that similar to other STAP detectors, the parametric Rao test is *adaptive* in that the detector is *data-dependent*, as evident in (7.14)-(3.21), which is in contrast to *data-independent* detector (e.g., a correlator). This shall not be confused with *recursive adaptive implementation*. Although recursive adaptive implementation of the parametric Rao test would be of interest in a real-time system, it is beyond the scope of the current section.

### 3.3.2 Asymptotic Analysis

As shown in Appendix A.3, the asymptotic distribution of the Rao/PAMF test statistic is given by

$$T_{\text{Rao}} \stackrel{a}{\sim} \begin{cases} \chi_2^2, & \text{under } H_0, \\ \chi_2'^2(\lambda), & \text{under } H_1, \end{cases} \quad (3.22)$$

where  $\chi_2^2$  denotes the central Chi-squared distribution with 2 degrees of freedom and  $\chi_2'^2(\lambda)$  the non-central Chi-squared distribution with 2 degrees of freedom and non-centrality parameter  $\lambda$ :

$$\lambda = 2|\alpha|^2 \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n), \quad (3.23)$$

where  $\tilde{\mathbf{s}}(n)$  is the temporally whitened steering vector given by (3.9). Note that  $\lambda$  is related to the SINR at the output of the temporal whitening filter. Recall that a  $\chi_2^2$  random variable is equivalent to an exponential random variable with probability density function (PDF) given by

$$f_{\chi_2^2}(x) = \frac{1}{2} \exp\left(-\frac{1}{2}x\right), \quad x \geq 0. \quad (3.24)$$

---

<sup>3</sup>The parametric GLRT is different from Kelly's GLRT in that the former takes into account the parametric model in Section 7.2, while the latter does not.

The PDF of  $\chi_2'^2(\lambda)$  is given by [7]

$$f_{\chi_2'^2(\lambda)}(x) = \frac{1}{2} \exp \left[ -\frac{1}{2} (x + \lambda) \right] I_0 \left( \sqrt{\lambda x} \right), \quad x \geq 0, \quad (3.25)$$

where  $I_0(u)$  is the modified Bessel function of the first kind and zero-th order defined by

$$I_0(u) = \frac{1}{\pi} \int_0^\pi \exp(u \cos \theta) d\theta = \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}u^2\right)^k}{(k!)^2}. \quad (3.26)$$

The above distributions can be employed to set the Rao test threshold for a given probability of false alarm, as well as to compute the detection and false alarm probabilities, and so on. For a given threshold, the probability of false alarm is given by

$$P_f = \int_{\gamma_{\text{Rao}}}^{\infty} f_{\chi_2^2}(x) dx = \exp \left( -\frac{1}{2} \gamma_{\text{Rao}} \right), \quad (3.27)$$

which can easily be inverted to find the test threshold  $\gamma_{\text{Rao}}$  for a given  $P_f$ . In addition, the probability of detection is given by

$$P_d = \int_{\gamma_{\text{Rao}}}^{\infty} \frac{1}{2} \exp \left[ -\frac{1}{2} (x + \lambda) \right] I_0 \left( \sqrt{\lambda x} \right) dx \quad (3.28)$$

for a given test threshold  $\gamma_{\text{Rao}}$ .

*Remark 5:* The asymptotic distribution under  $H_0$  is independent of the unknown parameters. The probability of false alarm in (3.27) depends only on the test threshold, which is a design parameter. It is evident that the Rao/PAMF test asymptotically achieves CFAR.

*Remark 6:* The above analysis holds under Assumptions **AS1** to **AS4** of Section 7.2 with one exception. In particular, since the ML parameter estimates are asymptotically Gaussian irrespective of the distribution of the observed data, the above analysis still holds if the Gaussian assumption in **AS3** is dropped. This also explains why it has been observed in several studies that the PAMF detector obtains good performance even with non-Gaussian observations (see, e.g., [22]).

### 3.4 Numerical Results

In the following, we present our numerical results of the parametric Rao/PAMF detector obtained by computer simulation and by the above asymptotic analysis. In addition, the performance of the MF (8.3) and AMF (8.5) detectors, which can be computed analytically, is included for comparison. For easy reference, Appendix A.4 contains a brief summary of relevant results, which have been used to compute the performance of the two detectors. We reiterate that the MF detector serves as a baseline only. We do not consider Kelly's GLRT since a detailed comparison between the GLRT and AMF detectors can be found in [12]. In the following, the disturbance

signal is generated as a multichannel AR(2) process with randomly generated AR coefficients  $\mathbf{A}$  and a spatial covariance matrix  $\mathbf{Q}$ . In particular,  $\mathbf{A}$  and  $\mathbf{Q}$  are selected to ensure that  $\mathbf{Q}$  is a valid covariance matrix and, furthermore,  $\mathbf{A}$  is chosen to ensure that the resulting AR process is stable. Once  $\mathbf{A}$  and  $\mathbf{Q}$  are selected, they are fixed in all trials. The signal vector  $\mathbf{s}$  is generated as in (3.4) with randomly chosen normalized spatial and Doppler frequencies. The SINR is defined as

$$\text{SINR} = |\alpha|^2 \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}, \quad (3.29)$$

where  $\mathbf{R}$  is the  $JN \times JN$  joint space-time covariance matrix of the disturbance  $\mathbf{d}$ , which can be determined once  $\mathbf{A}$  and  $\mathbf{Q}$  are selected (the details are not shown for simplicity). To numerically set the threshold for the parametric Rao/PAMF detector, a total of  $5 \times 10^4$  trials are run. Meanwhile, to determine  $P_d$  for a given threshold, a total of  $10^4$  trials are run for each SINR.

First, we consider the asymptotic distribution of the parametric Rao/PAMF test statistic obtained in Section 3.3.2. Figure 8.1 depicts the quantile-quantile plot of the Rao/PAMF test statistic under both hypotheses against the corresponding asymptotic distribution when  $J = 4$ ,  $N = 32$ , and  $K = 8$ , a case with limited training. It is seen that even with a relatively small data size, the asymptotic distribution matches well the sample test statistics, with only some minor deviation at the tail portion.

Next, we examine the *receiver operating characteristic* (ROC) [7] of the parametric Rao/PAMF detector. The parameters used in the simulation are  $J = 4$ ,  $N = 32$ , and  $K = 256$ . Figure 8.2 depicts the ROC curves for the parametric Rao/PAMF test obtained by simulation and asymptotic analysis, for SINR values of 0, 5, and 10 dB. It is seen that the simulation results match those obtained by asymptotic analysis.

Figures 8.3 to 8.6 depict the probability of detection  $P_d$  versus SINR for the MF, AMF, and the parametric Rao/PAMF detectors under various conditions that are specified below the figures. In particular, Figures 8.3 and 8.5 correspond to the case with adequate training, for which the RMB rule is satisfied (see discussions in Section 2), whereas Figures 8.4 and 8.6 correspond to the case with limited training, for which the AMF detector does not even exist, since the training size  $K = 8$  is too small to meet the minimum training condition (2.13). An examination of these figures reveals the following:

- When the assumptions of Section 7.2 are met, the asymptotic analysis provides a quite accurate prediction of the performance of the parametric Rao/PAMF detectors. The gap between the asymptotic and simulated results is seen to widen as  $K$  and/or  $N$  decreases. But even for the most challenging case with  $K = 8$  and  $N = 16$ , the gap is about 0.5 dB, as shown in Figure 8.6.
- The parametric Rao/PAMF detector is very close to the optimum MF detector. The gap between the two detectors closes with increasing  $K$  and/or  $N$ .
- The parametric Rao/PAMF detector outperforms the AMF detector by 2 to 3 dB when the RMB rule is marginally satisfied. This agrees with earlier observations made in [23].

So far we have assumed that the model order  $P$  of the multichannel AR process is known (cf. Assumption **AS4**). As mentioned in *Remark 1* of Section 7.2, various model selection tech-



niques can be used to estimate  $P$ , and it is not unusual for these techniques to under- or over-estimate the model order by a small number (relative to the true model order  $P$ ) [35, 36]. Hence, it would be of interest to find out how the parametric Rao/PAMF detector performs when an inaccurate model order estimate is used. This is shown in Figure 8.7, where the performance of the Rao/PAMF detector using the true, an under-estimated, and an over-estimated model order is depicted. As we can see, using an inaccurate model order estimate degrades the detection performance, but the degradation is not significant, especially in the case of model order over-estimation. Over-estimation is a more robust error since the high-order coefficients can be estimated close to zero (providing that the size of the signals that can be used for estimation is large enough). The above behavior of the parametric Rao/PAMF detector is typical and has been consistently observed in other experiments with a similar setup. Here, we only considered the case where the model order is incorrectly estimated by one unit. A larger performance variation is expected if there is a larger estimation error for  $P$ .

Finally, Figure 8.8 depicts  $P_d$  versus SINR for the parametric Rao/PAMF detector when  $J = 4$ ,  $P_f = 0.01$  and  $N$  varies from  $N = 4$  to  $N = 128$ . It is seen that the detection performance increases with  $N$ .

### 3.5 Conclusions

We have developed a parametric Rao test for the multichannel adaptive signal detection problem by exploiting a multichannel AR model. We have derived the ML estimates of the parameters involved in the test. The parametric Rao test is an asymptotic parametric GLRT, and the asymptotic distributions of its test statistic under both hypotheses have been obtained in closed form. We have shown that the PAMF test statistic has a form identical to that of the parametric Rao test statistic; therefore, the PAMF test is also an asymptotic parametric GLRT. Computer simulations show that: 1) our asymptotic analysis provides fairly accurate prediction of the performance of the parametric Rao/PAMF test; 2) even with relatively limited training, the parametric Rao/PAMF detector is quite close to the ideal MF detector; 3) the parametric Rao/PAMF detector outperforms the AMF detector, which does not exploit a parametric model; and finally 4) the performance of the parametric Rao/PAMF detector is affected by inaccurate model order estimation, but the resulting performance degradation is tolerable when the model order estimation error is small.

Our asymptotic analysis of the parametric Rao detector is based on several assumptions stated in Section 7.2, including that the disturbance can be modeled as an  $AR(P)$  process with known model order  $P$  and that the training signals are i.i.d. When these assumptions are violated, we expect that the analysis will be less accurate, but may be still informative if the assumptions are not significantly violated. For example, we have noticed in simulation that when the disturbance is an MA process, the test threshold obtained by analysis assuming an AR model is still quite accurate. One possible reason is that AR models are fairly general parametric models, and under mild conditions, can be used to model or approximate a large class of stationary random processes (e.g., an MA process can be approximated as an AR process with a high enough model order)

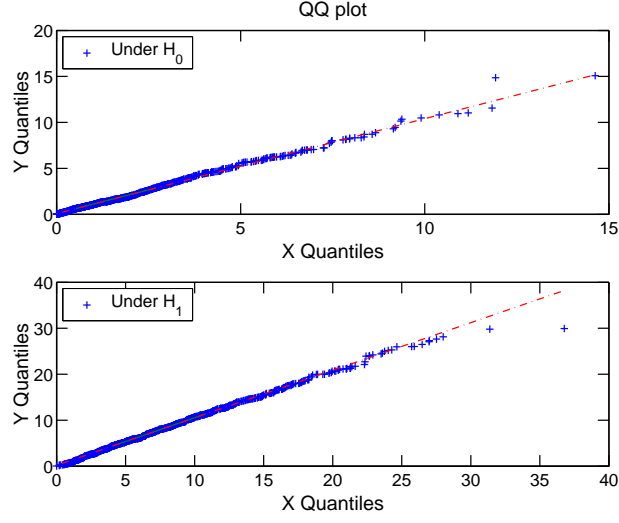


Figure 3.1: The quantile-quantile plot of the parametric Rao/PAMF test statistic and its asymptotic distribution under  $H_0$  (upper plot) and  $H_1$  (lower plot), respectively, with  $J = 4$ ,  $N = 32$ , and  $K = 8$ . Specifically, the x-axis shows the ordered samples of the parametric Rao/PAMF test statistic, while the y-axis shows the ordered samples of the asymptotic distribution.

[30]. Nevertheless, there is a need to find out how accurate our analysis is in real systems with real data when the assumptions of Section 7.2 may not all be met. This will be an interesting future effort.

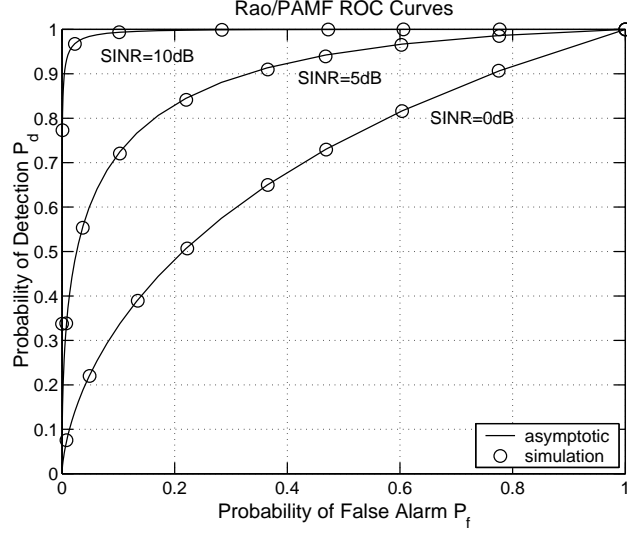


Figure 3.2: The receiver operating characteristics (ROC) curves of the parametric Rao/PAMF detector at various input SINR when  $J = 4$ ,  $N = 32$ , and  $K = 256$ .

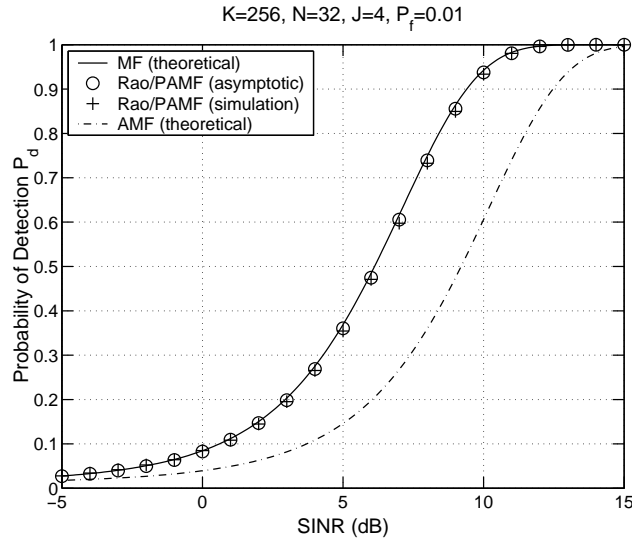


Figure 3.3: The probability of detection  $P_d$  versus the input SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 32$ , and  $K = 256$ .

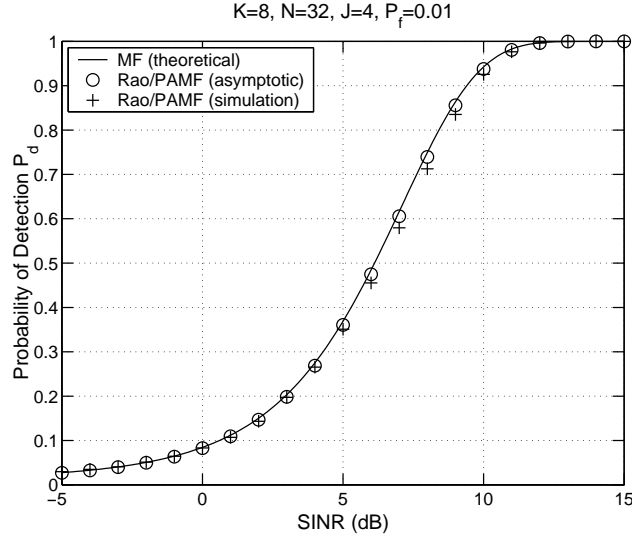


Figure 3.4: The probability of detection  $P_d$  versus the input SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 32$ , and  $K = 8$ . Note that the AMF detector is not included since it cannot be implemented for such a small  $K$ .

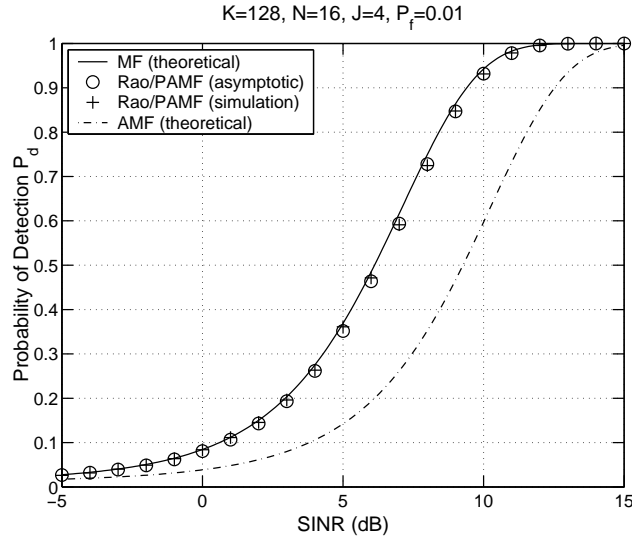


Figure 3.5: The probability of detection  $P_d$  versus the input SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 16$ , and  $K = 128$ .

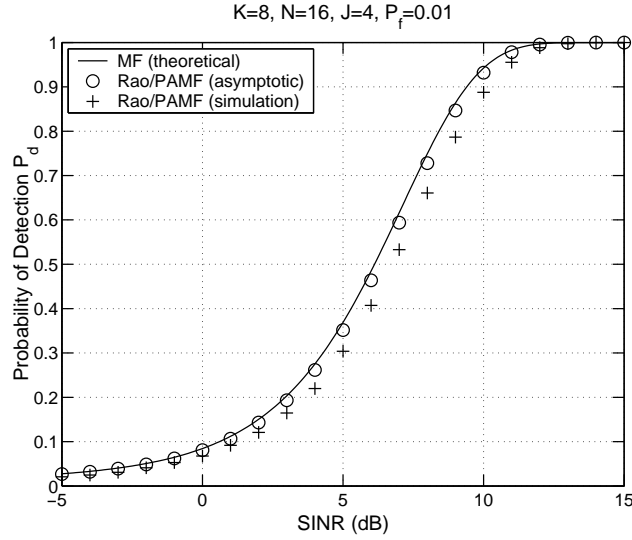


Figure 3.6: The probability of detection  $P_d$  versus the input SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 16$ , and  $K = 8$ . Note that the AMF detector is not included since it cannot be implemented for such a small  $K$ .

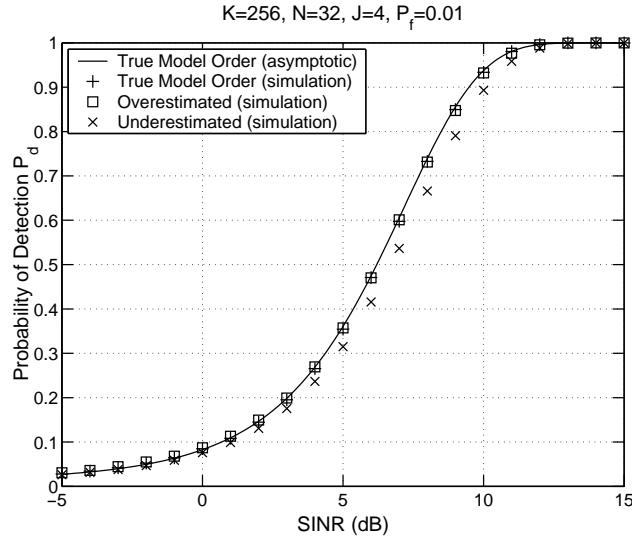
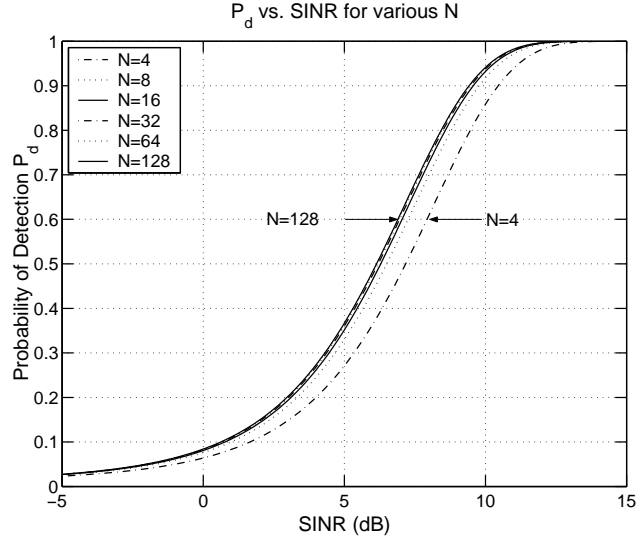
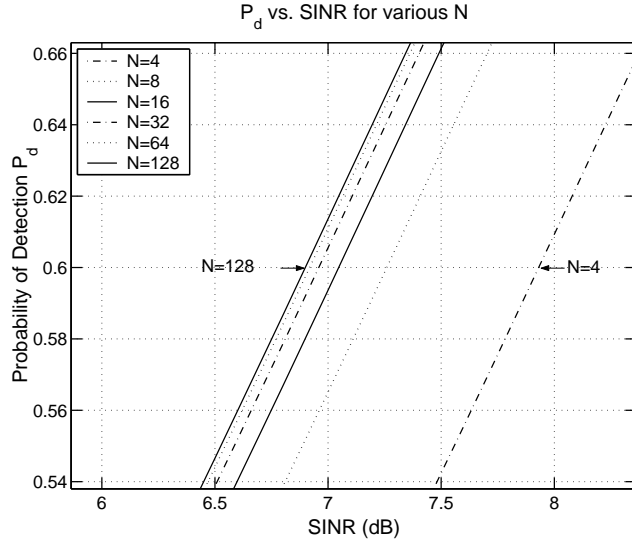


Figure 3.7: The probability of detection  $P_d$  versus the input SINR of the parametric Rao/PAMF detector when the model order of the multichannel AR process used for computing the test statistic is true ( $P = 2$ ), under-estimated (assuming  $P = 1$ ), and over-estimated (assuming  $P = 3$ ), along with  $P_f = 0.01$ ,  $J = 4$ ,  $N = 32$ , and  $K = 256$ .



(a)



(b)

Figure 3.8: (a) The impact of the pulse number  $N$  on the parametric Rao/PAMF detector when  $P_f = 0.01$  and  $J = 4$ . (b) “Zoomed-in” version of (a).

# Chapter 4

## Parametric GLRT for Multichannel Adaptive Signal Detection

### 4.1 Introduction

In this chapter, we develop a *parametric GLRT*. It is natural to extend the results of [37, 38] and consider the parametric GLRT for several reasons. First, as shown in Section 4.2, the problem of interest is a two-sided parameter testing problem that admits no uniformly most powerful (UMP) solution [7]. A GLRT approach is widely used in such cases due to its good asymptotic properties including asymptotic CFAR and consistency. Second, the parametric GLRT may yield improved performance than the parametric Rao detector, especially when the data is limited. This is because the latter is an asymptotic (large-sample) parametric GLRT [7, Appendix 6B]. Third, all Rao tests, including the parametric Rao detector, are obtained based on a further approximation that is valid only for weak signals [7, p. 238]. As such, the parametric Rao detector is expected to degrade when the weak signal assumption is violated. The above observations motivate us to consider the parametric GLRT, in hope of finding a better solution to the problem.

The parametric GLRT to be discussed is different from Kelly's GLRT [13]. The latter does not utilize a parametric model to model the disturbance. For this reason, our solution is referred to as the *parametric GLRT*. Our parametric GLRT is also different from the GLRT developed in [24], where a different detection problem is addressed that involves unknown non-linear signal parameters associated with the signal to be detected. We follow the direction of [9–13, 16, 22, 23] and consider a detection problem whereby the signal to be detected is known up to an unknown amplitude. The data model and assumptions for this problem are further discussed in Section 4.2.

The parametric GLRT relies on maximum likelihood (ML) parameter estimation for both the null and alternative hypotheses. The null hypothesis estimation problem is addressed in [37], where the ML estimator is obtained in closed-form. We show in Section 4.3.1 that the ML estimator under the alternative hypothesis is non-linear and requires searches on a two-dimensional parameter space. To address this issue, we introduce an asymptotic ML (AML) estimator that is considerably simpler, yielding estimates in a non-iterative fashion, and asymptotically coin-

cides with the optimum ML estimator. The AML estimator is related to an iterative alternating least-squares (ALS) estimator developed in [39], but with several notable distinctions (see Section 4.3.3). The Cramér-Rao bound (CRB) for the estimation problem is also derived, offering a baseline for comparing various (unbiased) estimators.

To examine the performance of the parametric GLRT, we consider scenarios with *very limited or even no training signals*. The less challenging case with more training is extensively considered in, e.g., [23, 37, 38], for the PAMF and parametric Rao detectors, which are equivalent to the parametric GLRT with a large amount of training. It should be noted that the parametric GLRT and Rao detectors utilize both test and training signals for parameter estimation; as such, they are functional even without training. *The capability to handle the training-free case is a unique and desirable attribute of the parametric GLRT and Rao detectors.* Although the performance of the parametric GLRT and Rao detector degrades in the absence of training, such degradation can be remedied by using a larger  $N$ , i.e., increasing temporal observations of the test signal. We show that the parametric GLRT outperforms the parametric Rao detector when  $N$  is small and, overall, the former yields a better detection performance.

The rest of the chapter is organized as follows. Section 4.2 contains the problem statement. Parameter estimation is addressed in Section 4.3, including the ML estimators for both hypotheses, the AML estimator for the alternative hypothesis, and the CRB. The test statistic, implementation, and asymptotic analysis for the parametric GLRT are discussed in Section 4.4. Numerical results are presented in Section 4.5, followed by our conclusions in Section 4.6.

## 4.2 Data Model and Problem Statement

In this section, we re-iterate the data model and problem statement for the completeness and easy reference. The problem of interest is to detect a known multichannel signal with unknown amplitude in the presence of spatially and temporally correlated disturbance (e.g., [1]):

$$\begin{aligned} H_0 : \quad & \mathbf{x}_0(n) = \mathbf{d}_0(n), \quad n = 0, 1, \dots, N-1, \\ H_1 : \quad & \mathbf{x}_0(n) = \alpha \mathbf{s}(n) + \mathbf{d}_0(n), \quad n = 0, 1, \dots, N-1, \end{aligned} \quad (4.1)$$

where all vectors are  $J \times 1$  with  $J$  denoting the number of spatial channels, and  $N$  is the number of temporal observations. In the sequel,  $\mathbf{x}_0(n)$  is referred to as the *test signal*,  $\mathbf{s}(n)$  is the signal to be detected with amplitude  $\alpha$ , and  $\mathbf{d}_0(n)$  is the *disturbance signal* that may be correlated in space and time. In addition to the test signal  $\mathbf{x}_0(n)$ , there may be a set of target-free *training* or *secondary* signals  $\mathbf{x}_k(n)$ ,  $k = 1, 2, \dots, K$ , to assist in the signal detection process:

$$\mathbf{x}_k(n) = \mathbf{d}_k(n), \quad n = 0, 1, \dots, N-1. \quad (4.2)$$

In radar systems, training data may be obtained from range cells adjacent to the test cell. However, training data is generally limited or may even be unavailable. In the training-free case, we have  $K = 0$ .



Define the following  $JN \times 1$  space-time vectors:

$$\begin{aligned} \mathbf{s} &= [\mathbf{s}^T(0), \mathbf{s}^T(1), \dots, \mathbf{s}^T(N-1)]^T, \\ \mathbf{d}_k &= [\mathbf{d}_k^T(0), \mathbf{d}_k^T(1), \dots, \mathbf{d}_k^T(N-1)]^T, \\ \mathbf{x}_k &= [\mathbf{x}_k^T(0), \mathbf{x}_k^T(1), \dots, \mathbf{x}_k^T(N-1)]^T, \\ &\quad k = 0, 1, \dots, K. \end{aligned} \tag{4.3}$$

It follows that (4.1) can be more compactly written as

$$\begin{aligned} H_0 : \quad \mathbf{x}_0 &= \mathbf{d}_0, \\ H_1 : \quad \mathbf{x}_0 &= \alpha \mathbf{s} + \mathbf{d}_0. \end{aligned} \tag{4.4}$$

Clearly, the composite hypothesis testing problem (4.1) or (4.4) is a *two-sided parameter testing problem* that tests  $\alpha = 0$  against  $\alpha \neq 0$ . The above signal detection problem occurs in an airborne STAP radar system with  $J$  array channels and a coherent processing interval (CPI) of  $N$  pulse repetition intervals (PRIs). The disturbance  $\mathbf{d}_k$  consists of ground clutter, jamming, and thermal noise, while  $\mathbf{s}$  is the *target space-time steering vector* [23].

The general assumptions in the literature are [1, 9–13, 15, 16, 22, 23]:

- **AS1:** The signal vector  $\mathbf{s}$  is deterministic and *known* to the detector;
- **AS2:** The signal amplitude  $\alpha$  is complex-valued, deterministic, and *unknown*;
- **AS3:** The disturbance signals  $\{\mathbf{d}_k\}_{k=0}^K$  are independent and identically distributed (i.i.d.) with distribution  $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ , where  $\mathbf{R}$  is the *unknown* space-time covariance matrix.

While **AS1** to **AS3** are standard [1, 9–13, 15, 22, 23]: we follow a parametric approach as in [22, 23, 38, 40]:

- **AS4:** The disturbance signal  $\mathbf{d}_k(n)$ ,  $k = 0, 1, \dots, K$ , can be modeled as a  $J$ -channel AR( $P$ ) process with *known*<sup>1</sup> model order  $P$ :

$$\mathbf{d}_k(n) = - \sum_{p=1}^P \mathbf{A}^H(p) \mathbf{d}_k(n-p) + \boldsymbol{\varepsilon}_k(n), \tag{4.5}$$

where  $\{\mathbf{A}^H(p)\}_{p=1}^P$  denote the *unknown*  $J \times J$  AR coefficient matrices,  $\boldsymbol{\varepsilon}_k(n)$  denote the driving  $J$ -channel spatial noise vectors that are temporally white but spatially colored Gaussian noise:  $\boldsymbol{\varepsilon}_k(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ , where  $\mathbf{Q}$  denotes the *unknown*  $J \times J$  spatial covariance matrix.

---

<sup>1</sup>If  $P$  is unknown, it can be estimated using a variety of model order selection techniques (e.g., [35, Appendix C]).

The *problem of interest* is to develop a GLRT based on Assumptions **AS1** to **AS4** for the above composite hypothesis testing problem, using the test signal  $\mathbf{x}_0$  and training signals  $\{\mathbf{x}_k\}_{k=1}^K$  if any. The likelihood functions under both hypotheses are parameterized by the signal parameter  $\alpha$  as well as nuisance parameters  $\mathbf{Q}$  and  $\mathbf{A}$ , where

$$\mathbf{A}^H = [\mathbf{A}^H(1), \mathbf{A}^H(2), \dots, \mathbf{A}^H(P)] \in \mathbb{C}^{J \times JP}. \quad (4.6)$$

For simplicity, we write the likelihood functions as

$$f_i(\alpha, \mathbf{A}, \mathbf{Q}), \quad i = 0 \text{ or } 1, \quad (4.7)$$

where  $\alpha = 0$  under  $H_0$  (i.e.,  $i = 0$ ) and  $\alpha \neq 0$  under  $H_1$  ( $i = 1$ ), and the dependence on the test/training signals  $\{\mathbf{x}_k\}_{k=0}^K$  is omitted. While the test statistic of the GLRT is well known, which is given by the generalized likelihood ratio (GLR) [7]:

$$\text{GLR} = \frac{\max_{\alpha, \mathbf{A}, \mathbf{Q}} f_1(\alpha, \mathbf{A}, \mathbf{Q})}{\max_{\mathbf{A}, \mathbf{Q}} f_0(0, \mathbf{A}, \mathbf{Q})}, \quad (4.8)$$

finding the ML estimates of the unknown parameters is non-trivial. We first address the estimation problem before examining the GLR test statistic in more details.

## 4.3 Parameter Estimation

Parameter estimators required by the parametric GLRT as well as the CRB are developed in Appendices B.1-B.4. The main results are summarized below.

### 4.3.1 ML Estimation under $H_1$

The ML estimate of  $\alpha$  under  $H_1$  is given by (see Appendix B.1)

$$\hat{\alpha}_{\text{ML}} = \arg \min_{\alpha} \left| \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha) \right|, \quad (4.9)$$

where the correlation matrices conditioned on  $\alpha$  are given by

$$\begin{aligned} \hat{\mathbf{R}}_{xx}(\alpha) &= \sum_{n=P}^{N-1} [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)] [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)]^H \\ &\quad + \sum_{n=P}^{N-1} \sum_{k=1}^K \mathbf{x}_k(n) \mathbf{x}_k^H(n), \end{aligned} \quad (4.10)$$

$$\begin{aligned} \hat{\mathbf{R}}_{yy}(\alpha) &= \sum_{n=P}^{N-1} [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)] [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)]^H \\ &\quad + \sum_{n=P}^{N-1} \sum_{k=1}^K \mathbf{y}_k(n) \mathbf{y}_k^H(n), \end{aligned} \quad (4.11)$$

$$\begin{aligned}\hat{\mathbf{R}}_{yx}(\alpha) &= \sum_{n=P}^{N-1} [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)] [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)]^H \\ &\quad + \sum_{n=P}^{N-1} \sum_{k=1}^K \mathbf{y}_k(n) \mathbf{x}_k^H(n),\end{aligned}\tag{4.12}$$

with

$$\mathbf{y}_k(n) = [\mathbf{x}_k^T(n-1), \dots, \mathbf{x}_k^T(n-P)]^T \tag{4.13}$$

$$\mathbf{t}(n) = [\mathbf{s}^T(n-1), \dots, \mathbf{s}^T(n-P)]^T. \tag{4.14}$$

Once  $\hat{\alpha}_{\text{ML}}$  is available, the ML estimates of  $\mathbf{A}$  and  $\mathbf{Q}$  under  $H_1$  are obtained as

$$\hat{\mathbf{A}}_{\text{ML},1}^H = \hat{\mathbf{A}}^H(\alpha)|_{\alpha=\hat{\alpha}_{\text{ML}}} \tag{4.15}$$

$$\hat{\mathbf{Q}}_{\text{ML},1} = \hat{\mathbf{Q}}(\alpha)|_{\alpha=\hat{\alpha}_{\text{ML}}}, \tag{4.16}$$

where the conditional estimates are given by

$$\hat{\mathbf{A}}^H(\alpha) = -\hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha), \tag{4.17}$$

$$\hat{\mathbf{Q}}(\alpha) = \frac{1}{L} \left[ \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha) \right], \tag{4.18}$$

with

$$L = (K+1)(N-P). \tag{4.19}$$

*Remark 1:* Although statistically optimum,  $\hat{\alpha}_{\text{ML}}$  has no closed-form expression. The cost function (4.9) is highly non-linear. A brute-force exhaustive search over the two-dimensional parameter space (i.e., the real and imaginary part of  $\alpha$ ) is generally impractical. Alternatively, we can resort to Newton-like iterative non-linear searches, provided an initial estimate of  $\alpha$  is available. Hence, there is a need for suboptimum estimators with reduced computational complexity. One such suboptimum estimator is discussed in Section 4.3.3.

### 4.3.2 ML Estimation under $H_0$

This is a special case of the one addressed in Section 4.3.1 (see Appendix B.1). The ML estimates of  $\mathbf{A}$  and  $\mathbf{Q}$  under  $H_0$  are given by

$$\hat{\mathbf{A}}_{\text{ML},0}^H = -\hat{\mathbf{R}}_{yx,0}^H \hat{\mathbf{R}}_{yy,0}^{-1}, \tag{4.20}$$

$$\hat{\mathbf{Q}}_{\text{ML},0} = \frac{1}{L} \left( \hat{\mathbf{R}}_{xx,0} - \hat{\mathbf{R}}_{yx,0}^H \hat{\mathbf{R}}_{yy,0}^{-1} \hat{\mathbf{R}}_{yx,0} \right), \tag{4.21}$$

where the correlation matrices  $\hat{\mathbf{R}}_{xx,0}$ ,  $\hat{\mathbf{R}}_{yy,0}$ ,  $\hat{\mathbf{R}}_{yx,0}$  are obtained from (4.10)-(4.12), respectively, by setting  $\alpha = 0$ .

### 4.3.3 Asymptotic ML Estimation under $H_1$

We now introduce a computationally more efficient estimator that is asymptotically equivalent to the ML estimator. The estimator is henceforth referred to as the AML estimator. The idea is to replace  $\hat{\mathbf{R}}_{yx}^H(\alpha)\hat{\mathbf{R}}_{yy}^{-1}(\alpha)$  in the cost function of (4.9) by a statistically consistent estimate  $\hat{\mathbf{A}}_{\text{con}}$  (how to obtain such a consistent estimate is discussed next). The resulting cost function  $C_1(\alpha)$ , which can be shown to be asymptotically equivalent to the cost function of (4.9) (e.g., [31, 41]), can be written as

$$C_1(\alpha) = \left| \sum_{n=P}^{N-1} \left\{ \tilde{\mathbf{x}}_0(n; \hat{\mathbf{A}}_{\text{con}}) - \alpha \tilde{\mathbf{s}}(n; \hat{\mathbf{A}}_{\text{con}}) \right\} \cdot \left\{ \tilde{\mathbf{x}}_0(n; \hat{\mathbf{A}}_{\text{con}}) - \alpha \tilde{\mathbf{s}}(n; \hat{\mathbf{A}}_{\text{con}}) \right\}^H + \sum_{n=P}^{N-1} \sum_{k=1}^K \tilde{\mathbf{x}}_k(n; \hat{\mathbf{A}}_{\text{con}}) \tilde{\mathbf{x}}_k^H(n; \hat{\mathbf{A}}_{\text{con}}) \right|, \quad (4.22)$$

where  $\hat{\tilde{\mathbf{x}}}_k(n)$  and  $\hat{\tilde{\mathbf{s}}}(n)$  are the temporally whitened versions of  $\mathbf{x}_k(n)$  and  $\mathbf{s}(n)$ , respectively, using the consistent AR coefficient estimate  $\hat{\mathbf{A}}_{\text{con}}$ :

$$\tilde{\mathbf{x}}_k(n; \hat{\mathbf{A}}_{\text{con}}) = \mathbf{x}_k(n) + \hat{\mathbf{A}}_{\text{con}}^H \mathbf{y}_k(n), \quad (4.23)$$

$$\tilde{\mathbf{s}}(n; \hat{\mathbf{A}}_{\text{con}}) = \mathbf{s}(n) + \hat{\mathbf{A}}_{\text{con}}^H \mathbf{t}(n). \quad (4.24)$$

Note that the matrix inside the determinant of (4.22) is a quadratic form of  $\alpha$ . An asymptotic solution is obtained in Appendix B.2, which is given by

$$\hat{\alpha}_{\text{AML}} = \frac{\text{tr} \left( \hat{\mathbf{S}}^H \mathbf{\Psi}^{-1} \hat{\mathbf{X}}_0 \right)}{\text{tr} \left( \hat{\mathbf{S}}^H \mathbf{\Psi}^{-1} \hat{\mathbf{S}} \right)}, \quad (4.25)$$

where

$$\hat{\mathbf{S}} = [\tilde{\mathbf{s}}(P; \hat{\mathbf{A}}_{\text{con}}), \dots, \tilde{\mathbf{s}}(N-1; \hat{\mathbf{A}}_{\text{con}})], \quad (4.26)$$

$$\hat{\mathbf{X}}_k = [\tilde{\mathbf{x}}_k(P; \hat{\mathbf{A}}_{\text{con}}), \dots, \tilde{\mathbf{x}}_k(N-1; \hat{\mathbf{A}}_{\text{con}})], \quad (4.27)$$

$$\mathbf{\Psi} = \hat{\mathbf{X}}_0 \mathbf{P}^\perp \hat{\mathbf{X}}_0^H + \sum_{k=1}^K \hat{\mathbf{X}}_k \hat{\mathbf{X}}_k^H, \quad (4.28)$$

with  $\mathbf{P}^\perp$  denoting the projection matrix projecting to the orthogonal complement of the range of  $\hat{\mathbf{S}}^H$ :

$$\mathbf{P}^\perp = \mathbf{I} - \mathbf{P} = \mathbf{I} - \hat{\mathbf{S}}^H \left( \hat{\mathbf{S}}^H \right)^\dagger \quad (4.29)$$

The above AML estimator requires a consistent estimate  $\hat{\mathbf{A}}_{\text{con}}$ , which can be obtained by using a consistent estimate of  $\alpha$  in (4.17). One such estimate is obtained by the least-squares

(LS) amplitude estimator:

$$\hat{\alpha}_{\text{LS}} = \frac{\mathbf{s}^H \mathbf{x}_0}{\mathbf{s}^H \mathbf{s}}, \quad (4.30)$$

which ignores the fact that the disturbance signal is colored. We show in Appendix B.3 that  $\hat{\alpha}_{\text{LS}}$  is statistically unbiased and consistent.

To summarize, the AML estimator can be implemented as follows:

- **Step 1:** Determine a consistent estimate  $\hat{\mathbf{A}}_{\text{con}}$ . This can be obtained by first computing the LS amplitude estimate  $\hat{\alpha}_{\text{LS}}$  as in (4.30), and using  $\hat{\alpha}_{\text{LS}}$  in (4.17):

$$\hat{\mathbf{A}}_{\text{con}}^H = -\hat{\mathbf{R}}_{yx}^H(\hat{\alpha}_{\text{LS}}) \hat{\mathbf{R}}_{yy}^{-1}(\hat{\alpha}_{\text{LS}}). \quad (4.31)$$

- **Step 2:** Compute the AML amplitude estimate  $\hat{\alpha}_{\text{AML}}$  using (4.25).
- **Step 3:** Find the AML estimates of the AR coefficients and spatial covariance matrix by substituting  $\hat{\alpha}_{\text{AML}}$  for  $\alpha$  in (4.17) and (4.18), respectively.

*Remark 2:* The AML estimator is obtained based on multiple approximations. The first involves approximating the likelihood function by dropping out the initial samples of the AR process, as shown in Appendix B.1. The second approximates the cost function in (4.9) by  $C_1(\alpha)$  as in (4.22), which replaces  $\mathbf{A}$  with a consistent estimate  $\hat{\mathbf{A}}_{\text{con}}$ .  $C_1(\alpha)$  is further shown to be equivalent to  $C_2(\alpha)$  in Appendix B.2. The third approximation is to replace the nonlinear  $C_2(\alpha)$  by a quadratic  $C_3(\alpha)$  in Appendix B.2, which admits a closed form solution. All three approximations are valid in the large-sample case.

*Remark 3:* The above AML estimator is related to an alternating LS (ALS) estimator discussed in [39], but there are several notable differences. First, AML covers both training ( $K \neq 0$ ) and training-free ( $K = 0$ ) cases, whereas ALS, which was introduced to solve an explosive detection problem, considers only the case without training. Second, ALS is an iterative approach, whereas iteration is not required by AML. Finally, by using an asymptotic approximation of the ML cost function, AML is directly related to the ML estimator and asymptotically coincides with the latter. Such an asymptotic relation was not established for ALS. Numerical results in Section 4.5 indicate that AML and ML yield nearly identical estimation performance.

### 4.3.4 CRB

From Section 4.3.1, an amplitude estimate is obtained first and then used to produce the nuisance parameter estimates. As such, amplitude estimation is the most critical step in the estimation process. Next, we provide the CRB for amplitude estimation. The CRB specifies a lower bound on the variance of any unbiased amplitude estimator, thus offering a baseline for comparison. The CRB for  $\alpha$  is derived in Appendix B.4, which is given by

$$\text{CRB}(\alpha) = \left[ \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n; \mathbf{A}) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n; \mathbf{A}) \right]^{-1}. \quad (4.32)$$

Like  $\hat{\mathbf{s}}(n)$  in (4.24),  $\tilde{\mathbf{s}}(n; \mathbf{A})$  is the temporally whitened version of  $\mathbf{s}(n)$ , but by using the true AR coefficient matrix  $\mathbf{A}$  (the dependence on  $\mathbf{A}$  is explicitly shown):

$$\tilde{\mathbf{s}}(n; \mathbf{A}) = \mathbf{s}(n) + \sum_{p=1}^P \mathbf{A}^H(p) \mathbf{s}(n-p). \quad (4.33)$$

The CRB for the nuisance parameters can be obtained in a similar fashion, but skipped for brevity.

## 4.4 Parametric GLRT

### 4.4.1 Test Statistic

With the ML parameter estimates obtained in Sections 4.3.1 and 4.3.2, the GLR reduces to

$$\text{GLR} = \frac{\max_{\alpha, \mathbf{A}, \mathbf{Q}} f_1(\alpha, \mathbf{A}, \mathbf{Q})}{\max_{\mathbf{A}, \mathbf{Q}} f_0(0, \mathbf{A}, \mathbf{Q})} = \frac{|\hat{\mathbf{Q}}_{\text{ML},0}|^L}{|\hat{\mathbf{Q}}_{\text{ML},1}|^L}, \quad (4.34)$$

where  $\hat{\mathbf{Q}}_{\text{ML},0}$  and  $\hat{\mathbf{Q}}_{\text{ML},1}$  are given by (4.21) and (4.16), respectively. Equivalently, taking a logarithm (with a scaling constant 2) yields<sup>2</sup>

$$T_{\text{GLRT}} = 2L \ln \left| \frac{\hat{\mathbf{Q}}_{\text{ML},0}}{\hat{\mathbf{Q}}_{\text{ML},1}} \right| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{GLRT}}, \quad (4.35)$$

where  $\gamma_{\text{GLRT}}$  denotes the test threshold (see Section 4.4.2 for discussion on the setting of  $\gamma_{\text{GLRT}}$ ).

*Remark 4:* The final test statistic is a ratio of two matrix determinants. Note that the two covariance matrix estimates  $\hat{\mathbf{Q}}_{\text{ML},0}$  and  $\hat{\mathbf{Q}}_{\text{ML},1}$  have an identical form given by (4.17), except that  $\alpha = 0$  for the former and  $\alpha = \hat{\alpha}_{\text{ML}}$  for the latter. Hence, once  $\hat{\alpha}_{\text{ML}}$  is obtained, the remaining steps involved in calculating  $\hat{\mathbf{Q}}_{\text{ML},1}$  are very similar to those needed for  $\hat{\mathbf{Q}}_{\text{ML},0}$ , which can be performed by the same computing algorithm or hardware, thus simplifying implementation.

*Remark 5:* Instead of using the non-linear ML amplitude estimate  $\hat{\alpha}_{\text{ML}}$ , we can employ the computationally more efficient AML amplitude estimate  $\hat{\alpha}_{\text{AML}}$  in calculating the GLRT test statistic. As shown in Section 4.5, the two different versions of GLRT offer nearly identical detection performance.

---

<sup>2</sup>Although the factor of  $2L$  and logarithm in the test statistic can be absorbed by the test threshold, it is retained to keep the asymptotic distribution more compact. See Section 4.4.2.

### 4.4.2 Asymptotic Analysis

As shown in Appedix B.5, the asymptotic distribution of the parametric GLRT statistic in (7.6) is given by

$$T_{\text{GLRT}} \stackrel{a}{\sim} \begin{cases} \chi_2^2, & \text{under } H_0, \\ \chi_2'^2(\lambda), & \text{under } H_1, \end{cases} \quad (4.36)$$

where  $\chi_2^2$  denotes the central Chi-squared distribution with 2 degrees of freedom (i.e., exponential distribution) and  $\chi_2'^2(\lambda)$  the non-central Chi-squared distribution with 2 degrees of freedom and non-centrality parameter  $\lambda$ :

$$\lambda = 2|\alpha|^2 \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n; \mathbf{A}) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n; \mathbf{A}). \quad (4.37)$$

Note that  $\lambda$  is related to the signal-to-interference-plus-noise ratio (SINR) at the output of the temporal whitening filter. Using the above result, we can write the the asymptotic detection and false alarm probabilities as

$$P_d = \int_{\gamma_{\text{GLRT}}}^{\infty} \frac{1}{2} \exp \left[ -\frac{1}{2} (x + \lambda) \right] I_0 \left( \sqrt{\lambda x} \right) dx, \quad (4.38)$$

$$P_f = \exp \left( -\frac{1}{2} \gamma_{\text{GLRT}} \right), \quad (4.39)$$

where  $I_0(u)$  is the modified Bessel function of the first kind and zero-th order [7].

*Remark 6:* The asymptotic distribution under  $H_0$  is independent of the unknown parameters. The probability of false alarm in (4.39) depends only on the test threshold, which is a design parameter. It is evident that the parametric GLRT asymptotically achieves CFAR.

*Remark 7:* The above analysis holds under Assumptions **AS1** to **AS4** of Section 4.2 with one exception. In particular, since the ML parameter estimates are asymptotically Gaussian irrespective of the distribution of the observed data, the above analysis still holds if the Gaussian assumption in **AS3** is dropped.

## 4.5 Numerical Results

In this section, we present simulation results to illustrate the performance of the proposed detection and estimation techniques. The disturbance signal is generated as a multichannel AR(2) process with AR coefficients  $\mathbf{A}$  and a spatial covariance matrix  $\mathbf{Q}$ . These parameters are set to ensure that the AR process is stable and  $\mathbf{Q}$  is a valid covariance matrix, but otherwise randomly selected. The signal vector  $\mathbf{s}$  corresponds to a uniform equi-spaced linear array with  $J = 4$  antenna elements and randomly selected normalized spatial and Doppler frequencies (see [23]). The SINR is defined as

$$\text{SINR} = |\alpha|^2 \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}, \quad (4.40)$$

where the  $JN \times JN$  space-time covariance matrix can be uniquely determined once  $\mathbf{A}$  and  $\mathbf{Q}$  are selected. Note that the above SINR can be considered as an *overall* SINR that takes into account all spatial and temporal signals observed within one CPI. A different SINR that is also frequently used is defined based on one snapshot of the array output; see, e.g., [39].

### 4.5.1 Estimation

The estimation results are presented for the estimators discussed in Section 4.3, namely, the LS (4.30), AML (4.25), and ML (4.9) estimators. The ML estimator is implemented via local non-linear iterative searching, initialized by the AML estimate. We first consider the *training-free* case with  $K = 0$ . This is also the case considered by the ALS estimator [39] and, thus, we include it for comparison. Figure 4.1 shows the mean-squared error (MSE) of the amplitude estimate  $\hat{\alpha}$  obtained by each estimator, along with the CRB (4.32), versus the SINR. It is seen that even for a moderate value of  $N = 32$ , the AML amplitude estimate is nearly identical to the ML estimate, and both are very close to the CRB and considerably better than the simple LS estimate. It is also observed that the ALS estimate is nearly identical to the AML estimate in this case.

Figure 4.2 depicts the results for a *limited-training* case with  $K = 1$ . The LS and ALS estimators are not included since they do not utilize any training signal for estimation. It is seen that both the AML and ML estimates are nearly identical and close to the CRB for all values of SINR. It is observed that use of training data slightly improves the estimation performance.

### 4.5.2 Detection

For the parametric GLRT (7.6), the test statistic can be computed using either the ML or AML parameter estimates, as indicated in Remark 5 of Section 4.4.1. The resulting tests, which are denoted as *parametric GLRT/ML* and *parametric GLRT/AML*, respectively, are compared with the parametric Rao detector [37, 38], which is a large-sample approximation of the parametric GLRT. Also included in the comparison are the asymptotic analysis for the parametric GLRT given in Section 4.4.2, the ideal matched filter (MF) [12] which assumes exact knowledge of  $\mathbf{R}$  and, therefore, cannot be used in practice but offers a baseline for comparison, and Kelly's GLRT [13] which is included to show the gain offered by parametric detection.<sup>3</sup> In all examples, we set  $J = 4$  and the probability of false alarm  $P_f = 0.01$ .

The *training-free* ( $K = 0$ ) case is considered in Figures 4.3 to 4.5, which show the probability of detection versus SINR for various detectors when the number of temporal observations varies from  $N = 32, 64$ , to 128. Meanwhile, the *limited-training* ( $K = 1$ ) case is considered in Figures 4.6 and 4.7 for  $N = 32$  and 64, respectively. An examination of these figures reveals the following:

---

<sup>3</sup>Recall that Kelly's GLRT is a covariance matrix based detector that cannot handle the limited-training or training-free case. In the following examples, we use either  $K = 0$  or  $K = 1$  for the parametric detectors; but for Kelly's GLRT,  $K$  is chosen significantly larger so that  $K \geq JN - 1$  to ensure a non-singular estimate of  $\mathbf{R}$  (see discussions in Section 4.1).



- The parametric GLRT/AML yields nearly identical detection performance to that of the parametric GLRT/ML, and may be preferred to the latter due to its reduced complexity.
- For the training-free case, the parametric GLRT is about 3 to 4 dB from the optimum MF bound at  $N = 32$ ; the gap reduces to about 1 dB at  $N = 64$  and a fractional dB at  $N = 128$ . Training, even modest, helps improving the detection, which can be seen by comparing Figure 4.3 with Figure 4.6, or Figure 4.4 with Figure 4.7. However, *the degradation incurred by lack of training can be remedied by increasing temporal observations of the test signal*, as seen in Figures 4.3 to 4.5.
- For small  $N$  (e.g.,  $N = 32$ ), the parametric GLRT outperforms the parametric Rao detector. At larger values of  $N$ , the two detectors exhibit similar performance, especially at the low SINR region.
- The parametric Rao detector is seen to degrade dramatically as the SINR increases. This is not surprising since all Rao tests, including the parametric Rao detector, are based on a *weak* signal approximation of the GLRT [7, Appendix 6B]. This has also been observed in [42, Fig. 3] for a single-channel detection problem. Such degradation may not be critical in applications where weak signal detection is of primary interest.
- Compared to Kelly's GLRT, both parametric detectors can produce better detection performance with significantly less training or even no training, when  $N$  is not too small.

## 4.6 Conclusions

We have developed a new parametric GLRT for multichannel adaptive signal detection. The parametric GLRT is obtained by exploiting multichannel AR modeling for the disturbance signal. We have investigated the underlying parameter estimation problem. The ML estimator has been derived, but not in a closed form. An AML estimator has been introduced as an asymptotically optimum but computationally more efficient alternative. We have examined the detection performance of the parametric GLRT as well as a recently proposed parametric Rao detector. We have shown that while both parametric detectors are significantly less dependent on training than conventional covariance matrix based detectors, the parametric GLRT is the better solution of the two, especially when temporal observations of the test signal are limited.

One most interesting feature of the parametric GLRT and Rao detectors is that both use the test and training signals for parameter estimation and can handle the training-free case. We have shown that the performance degradation caused by the lack of training can be remedied by increasing the temporal observations of the test signal. Such a tradeoff may be of interest and exploited in some applications, such as radars, when the environment is highly heterogeneous such that using neighboring range cells for training becomes impossible. In particular, the i.i.d. assumption **AS3** concerning the test cell and the neighboring range cells will be seriously violated in that case.

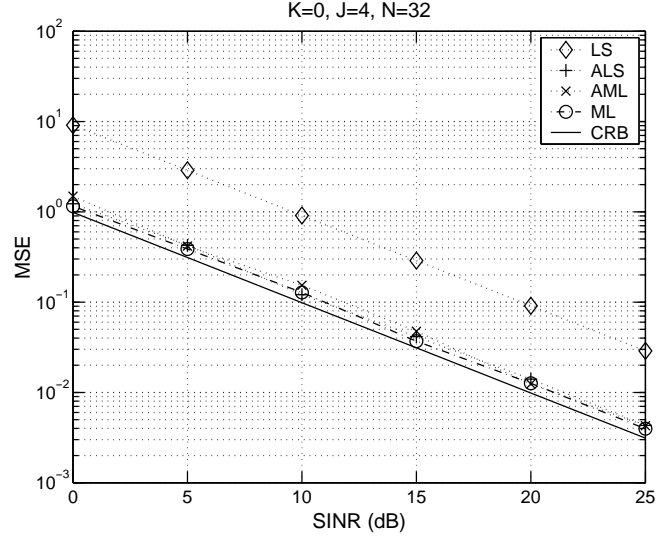


Figure 4.1: MSE of amplitude estimate  $\hat{\alpha}$  versus SINR when  $J = 4$ ,  $N = 32$ , and  $K = 0$  (no training data).

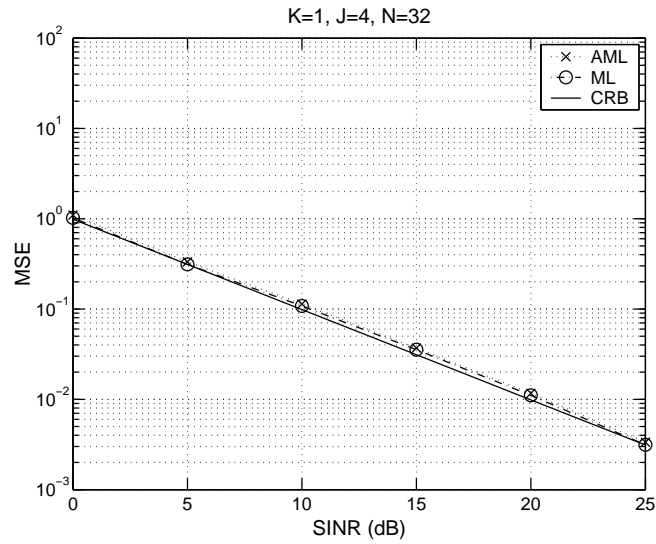


Figure 4.2: MSE of amplitude estimate  $\hat{\alpha}$  versus SINR when  $J = 4$ ,  $N = 32$ , and  $K = 1$  (limited training data).

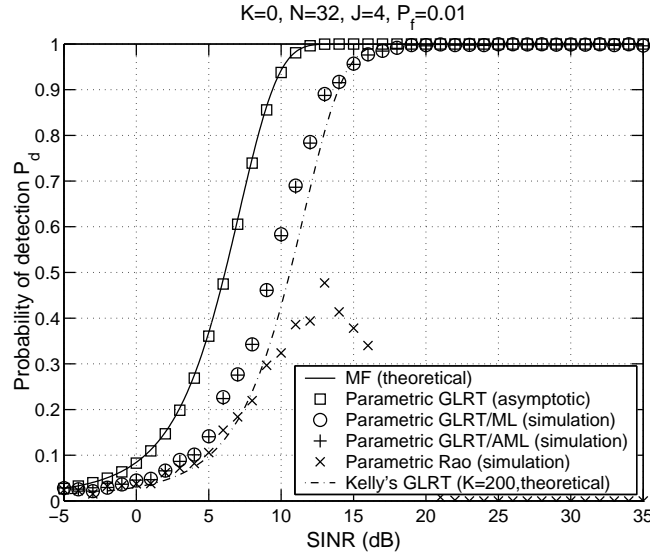


Figure 4.3: Probability of detection  $P_d$  versus SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 32$ , and  $K = 0$  (no training data).

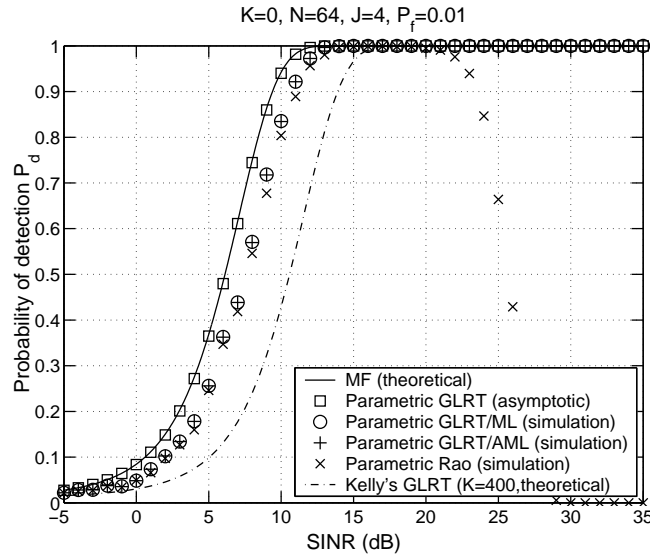


Figure 4.4: Probability of detection  $P_d$  versus SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 64$ , and  $K = 0$  (no training data).

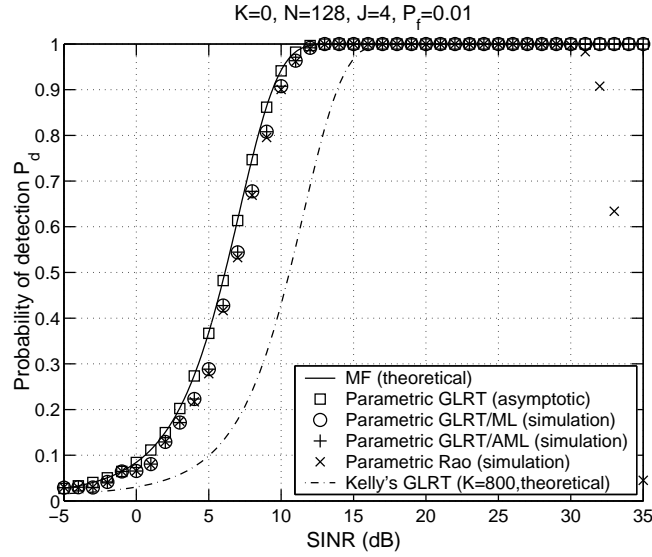


Figure 4.5: Probability of detection  $P_d$  versus SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 128$ , and  $K = 0$  (no training data).

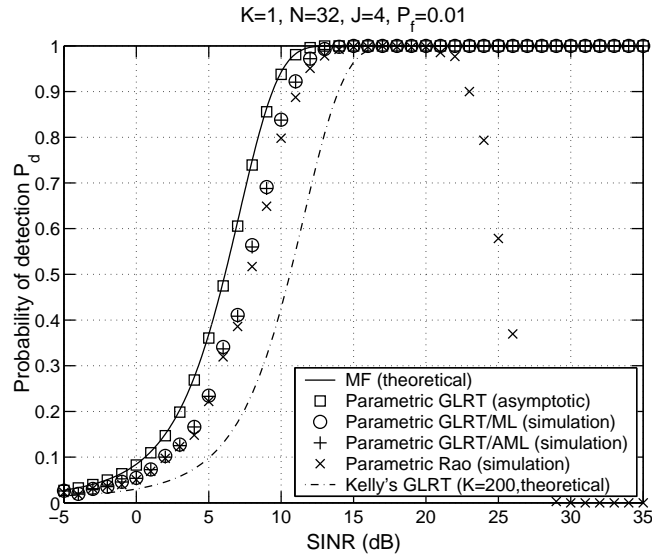


Figure 4.6: Probability of detection  $P_d$  versus SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 32$ , and  $K = 1$  (limited training data).

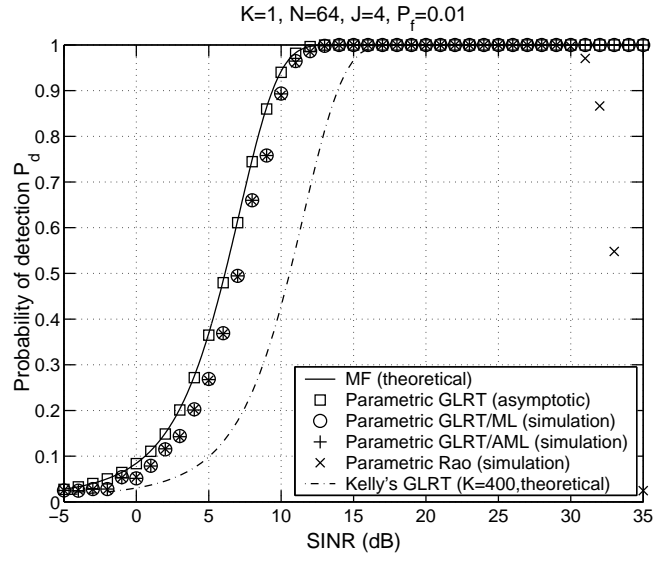


Figure 4.7: Probability of detection  $P_d$  versus SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 64$ , and  $K = 1$  (limited training data).

## Chapter 5

# A Simplified Parametric GLRT for Multichannel Adaptive Signal Detection

### 5.1 Introduction

Parametric STAP detectors have recently gained considerable interest due to their remarkable ability of offering significant performance improvement over classical detectors in training limited cases. Specifically, the parametric adaptive matched filter (PAMF) [22], one of the first in this class, models the disturbance signal as a parametric multichannel autoregressive (AR) process. The parametric model allows signal whitening through an inverse moving-average filter, which replaces the standard whitening process using a full-dimensional space-time covariance matrix estimate found in classical STAP detectors. The immediate benefit brought by the parametric model is reduced unknown parameters to be estimated and, in turn, reduced training and computational requirements. The multichannel AR process has been found to be an effective tool to model real-world airborne radar clutter for STAP detection [23, 43, 44]. It is also versatile in capturing the temporal and spatial correlation of disturbance signals in other radar and array processing applications (e.g., [24, 25, 39]). The PAMF detector is shown to be equivalent to a parametric Rao detector in [28]. The equivalence leads to analytical expressions for the asymptotic performance of the PAMF detector. Efficient implementations of the PAMF detector capitalizing on the inherent computational structure of the multichannel AR model are reported in [45, 46]. Meanwhile, extensions of the multichannel AR modeling to non-stationary cases for STAP detection are investigated in [47–50].

The parametric generalized likelihood ratio test (GLRT) [28] is a recent addition into the parametric STAP family. An interesting observation made in [28] is that it is possible to trade range training with the number of pulses within a coherent processing interval (CPI). Specifically, the traditional way of learning the clutter statistic is to use the signals received over adjacent range cells near the test range as the training signals, assuming that the target is a rare event and that the clutter statistic does not change much over the neighborhood of the cell under test. This assumption is clearly violated in heterogeneous dense-target environments, which is why the sample covariance matrix based detectors do not perform well in such cases. In contrast, [28]

shows that the clutter statistic can be extracted from the temporal pulses over a CPI; in the extreme case, this can be achieved exclusively from the test signal, without using any range training, provided that the number of pulses is large enough. The performance of the parametric GLRT has been examined using simulated and real data in various training limited cases [43, 44].

There are still critical unresolved issues with the parametric GLRT. Specifically, it involves highly nonlinear parameter estimation that has no closed-form solution. An iterative search based procedure is employed in [28], which is seen to be computationally intensive. Moreover, the iterative searching requires an initial guess of the unknown parameter. A two-step estimator is presented for that purpose, which starts with a least-squares (LS) estimation step by ignoring the temporal/spatial correlation, followed by a refining step. While this estimator is an asymptotic maximum likelihood (AML) estimator, its performance is limited by the coarse LS estimator and, as we show in Section 5.5, may not perform well when the number of pulses is small. Finally, the parametric GLRT, due to its complicated nonlinear form, offers little insight into how it functions. This is different from other parametric STAP detectors (e.g., [23, 40]) which have a clear interpretation of sequential temporal and spatial whitening (see discussions in Section 5.4 for details).

To address the above issues, we present herein a new estimator for the estimation problem underlying the parametric GLRT. The new estimator is in closed-form and computationally simple. Unlike the earlier AML estimator, it does not need an initial guess and, thus, is not hindered by poor initialization. The new estimator also leads to a simplified parametric GLRT, offering additional insight unavailable with the original GLRT. In general, the new GLRT achieves similar detection performance as the original one. But in the more challenging case when the number of pulses is limited, the new GLRT may outperform the original GLRT (which employs an iterative search based estimation procedure initialized by the AML estimator). The performance loss of the latter is primarily due to the poor initial parameter estimate provided by the AML estimator.

The remainder of this chapter is organized as follows. Section 5.2 contains the data model and a summary of the original GLRT of [28], where an underlying nonlinear amplitude estimation problem is also highlighted. A new amplitude estimator is introduced in Section 5.3, which leads to a simplified parametric GLRT presented in Section 5.4. Numerical results and conclusions are provided in Section 5.5 and 5.6, respectively.

## 5.2 Data Model and Parametric GLRT

### 5.2.1 Data Model

The problem of interest is to detect a  $JN \times 1$  multichannel signal  $\mathbf{s}$  with *unknown* amplitude  $\alpha$  in the presence of spatially and temporally correlated disturbance: (e.g., [1]):

$$\begin{aligned} H_0 : \quad \mathbf{x}_0 &= \mathbf{d}_0, \\ H_1 : \quad \mathbf{x}_0 &= \alpha \mathbf{s} + \mathbf{d}_0, \end{aligned} \tag{5.1}$$

where  $J$  denotes the number of spatial channels and  $N$  the number of temporal observations (i.e., snapshots). It will be convenient to express the  $JN \times 1$  vectors in terms of their spatial  $J \times 1$  components, i.e.,

$$\mathbf{s} = [\mathbf{s}^T(0), \dots, \mathbf{s}^T(N-1)]^T, \quad (5.2)$$

and similarly  $\mathbf{d}_0$  and  $\mathbf{x}_0$  are decomposed into  $\mathbf{d}_0(n)$  and  $\mathbf{x}_0(n)$ , respectively. In the sequel,  $\mathbf{x}_0(n)$  is referred to as the *test signal*,  $\mathbf{s}(n)$  as the *steering vector* (assumed known to the detector), and  $\mathbf{d}_0(n)$  as the *disturbance signal* (i.e., clutter and noise) that may be correlated in space and time. In addition to the test signal  $\mathbf{x}_0(n)$ , there may be a set of *training* or *secondary* signals  $\mathbf{x}_k(n)$ ,  $k = 1, \dots, K$ , that are target-free:  $\mathbf{x}_k(n) = \mathbf{d}_k(n)$ .

The binary composite hypothesis testing problem is to select between  $H_0 : \alpha = 0$  and  $H_1 : \alpha \neq 0$ . A standard assumption in STAP detection (e.g., [1, 9–13, 15, 16]) is that the disturbance signals  $\{\mathbf{d}_k\}_{k=0}^K$  are independent and identically distributed (i.i.d.) with distribution  $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ , where  $\mathbf{R}$  is the *unknown* space-time covariance matrix. The parametric STAP detectors [22, 23, 28, 40] further assume that the disturbance signal  $\mathbf{d}_k(n)$ ,  $k = 0, \dots, K$ , can be modeled as a  $J$ -channel AR( $P$ ) process:

$$\mathbf{d}_k(n) = -\sum_{i=1}^P \mathbf{A}^H(i) \mathbf{d}_k(n-i) + \boldsymbol{\varepsilon}_k(n), \quad (5.3)$$

where  $\{\mathbf{A}^H(i)\}_{i=1}^P$  denote the *unknown*  $J \times J$  AR coefficient matrices,  $\boldsymbol{\varepsilon}_k(n)$  denote the  $J \times 1$  spatial noise vectors that are assumed to be temporally white but spatially colored Gaussian noise:  $\boldsymbol{\varepsilon}_k(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ , where  $\mathbf{Q}$  denotes the *unknown*  $J \times J$  spatial covariance matrix.

### 5.2.2 Parametric GLRT

To introduce the necessary notation and also to facilitate comparison, we briefly summarize the parametric GLRT [28] as follows. The parametric GLRT first finds the ML estimates (MLEs) of the unknown parameters under both hypotheses, which are next used to compute the test statistics. Amplitude estimation under  $H_1$  turns out to be the key problem, as the other parameters can be readily obtained once an estimate of  $\alpha$  is available. Specifically, the MLE of  $\alpha$  is given by

$$\hat{\alpha}_{\text{ML}} = \arg \min_{\alpha} \left| \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha) \right|, \quad (5.4)$$



where  $\hat{\mathbf{R}}_{xx}(\alpha)$ ,  $\hat{\mathbf{R}}_{yy}(\alpha)$  and  $\hat{\mathbf{R}}_{yx}(\alpha)$  are  $J \times J$ ,  $JP \times JP$  and  $JP \times J$  matrices defined as

$$\begin{aligned} \hat{\mathbf{R}}_{xx}(\alpha) &= \sum_{k=1}^K \sum_{n=P}^{N-1} \mathbf{x}_k(n) \mathbf{x}_k^H(n) \\ &\quad + \sum_{n=P}^{N-1} [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)] [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)]^H, \end{aligned} \quad (5.5)$$

$$\begin{aligned} \hat{\mathbf{R}}_{yy}(\alpha) &= \sum_{k=1}^K \sum_{n=P}^{N-1} \mathbf{y}_k(n) \mathbf{y}_k^H(n) \\ &\quad + \sum_{n=P}^{N-1} [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)] [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)]^H, \end{aligned} \quad (5.6)$$

$$\begin{aligned} \hat{\mathbf{R}}_{yx}(\alpha) &= \sum_{k=1}^K \sum_{n=P}^{N-1} \mathbf{y}_k(n) \mathbf{x}_k^H(n) \\ &\quad + \sum_{n=P}^{N-1} [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)] [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)]^H, \end{aligned} \quad (5.7)$$

and the regression vectors are defined as  $\mathbf{t}(n) = [\mathbf{s}^T(n-1), \dots, \mathbf{s}^T(n-P)]^T \in \mathbb{C}^{JP \times 1}$  and  $\mathbf{y}_k(n) = [\mathbf{x}_k^T(n-1), \dots, \mathbf{x}_k^T(n-P)]^T \in \mathbb{C}^{JP \times 1}$ ,  $k = 0, \dots, K$ . Once  $\hat{\alpha}_{\text{ML}}$  is obtained, the parametric GLRT is given by

$$T_{\text{GLRT}} = 2L \ln \frac{|\hat{\mathbf{Q}}_{\text{ML},0}|}{|\hat{\mathbf{Q}}_{\text{ML},1}|} \underset{H_0}{\overset{H_1}{\geq}} \gamma_{\text{GLRT}}, \quad (5.8)$$

where<sup>1</sup>  $L = (K+1)(N-P)$ ,  $\gamma_{\text{GLRT}}$  denotes the corresponding test threshold, and  $\hat{\mathbf{Q}}_{\text{ML},0}$  and  $\hat{\mathbf{Q}}_{\text{ML},1}$  denote the ML estimates of the spatial covariance matrix under the null and alternative hypotheses

$$\hat{\mathbf{Q}}_{\text{ML},0} = \hat{\mathbf{Q}}(\alpha)|_{\alpha=0}, \quad (5.9)$$

$$\hat{\mathbf{Q}}_{\text{ML},1} = \hat{\mathbf{Q}}(\alpha)|_{\alpha=\hat{\alpha}_{\text{ML}}}, \quad (5.10)$$

with

$$\hat{\mathbf{Q}}(\alpha) = \frac{1}{L} \left( \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha) \right). \quad (5.11)$$

The MLE (7.13) is highly nonlinear and cannot be solved in closed-form. Iterative search over a two-dimensional (2D) parameter space (note that  $\alpha$  is complex-valued) is typically employed, which is computationally intensive (as the matrix determinant has to be evaluated for

---

<sup>1</sup>While the scaling factor  $L$  can be dropped from the test statistic, it was retained in [28] to simplify the asymptotic analysis.

every update of  $\alpha$ ) and, in general, converges only to a local minimum. To address this problem, an asymptotic ML (AML) estimator was introduced in [28]. The AML estimator involves a two-step process. In particular, it first computes the least-squares (LS) of the amplitude

$$\hat{\alpha}_{\text{LS}} = \frac{\mathbf{s}^H \mathbf{x}_0}{\mathbf{s}^H \mathbf{s}}, \quad (5.12)$$

which effectively ignores the spatio-temporal correlation of the disturbance signal. Then, the initial estimate is refined through a weighed LS process (see [28] for details). Although the AML estimate can be shown to be asymptotically efficient, it is affected by the limited performance of the initializing LS estimator, in particular when the data size is small.

In closing this section, we briefly comment on the stability issue. In general, the multichannel AR process used to model the disturbance signal has to be stable to ensure that the resulting AR signal is wide-sense stationary [30]. A constrained ML estimator that maximizes the likelihood function under the constraint that the AR coefficient matrices form a stable multichannel filter is highly involved and generally not suitable for practical applications. In contrast, the estimators considered in this work, including the ML and AML estimators as well as the one introduced in the next section, do not impose this constraint in seek of computational simplicity. Although the estimated AR model obtained by any of these estimators is not guaranteed to be stable, extensive numerical studies using simulated and experimental data show that these unconstrained estimators yield good estimation and detection performance at acceptable complexity.

### 5.3 Amplitude Estimation

The exact MLE (7.13) minimizes the determinant of the  $\alpha$ -dependent matrix

$$\hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha),$$

which is the Schur complement (see, e.g., [51]) of the block matrix  $\hat{\mathbf{R}}(\alpha)$

$$\hat{\mathbf{R}}(\alpha) = \begin{bmatrix} \hat{\mathbf{R}}_{yy}(\alpha) & \hat{\mathbf{R}}_{yx}(\alpha) \\ \hat{\mathbf{R}}_{yx}^H(\alpha) & \hat{\mathbf{R}}_{xx}(\alpha) \end{bmatrix}. \quad (5.13)$$

It is well-known that the determinant of a block matrix like (5.13) can be expressed in terms of its Schur complement [51]:

$$\left| \hat{\mathbf{R}}(\alpha) \right| = \left| \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha) \right| \left| \hat{\mathbf{R}}_{yy}(\alpha) \right|.$$

Using the above result, the cost function in (7.13) is equivalent to

$$\ln \left| \hat{\mathbf{R}}(\alpha) \right| - \ln \left| \hat{\mathbf{R}}_{yy}(\alpha) \right|. \quad (5.14)$$

By using (7.10)-(7.12), along with new definitions of regression vectors

$$\mathbf{s}_{P+1}(n) = [\mathbf{t}^T(n), \mathbf{s}^T(n)]^T, \quad (5.15)$$

$$\mathbf{x}_{k,P+1}(n) = [\mathbf{y}_k^T(n), \mathbf{x}_k^T(n)]^T, \quad (5.16)$$

$\hat{\mathbf{R}}(\alpha)$  can be decomposed into an  $\alpha$ -dependent component and an  $\alpha$ -independent one:

$$\hat{\mathbf{R}}(\alpha) = (\mathbf{X}_0 - \alpha \mathbf{S})(\mathbf{X}_0 - \alpha \mathbf{S})^H + \sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^H, \quad (5.17)$$

where the new steering matrix  $\mathbf{S}$  and data matrix  $\mathbf{X}_k$  are given as

$$\mathbf{S} = [\mathbf{s}_{P+1}(P), \dots, \mathbf{s}_{P+1}(N-1)] \in \mathbb{C}^{J(P+1) \times (N-P)}, \quad (5.18)$$

$$\mathbf{X}_k = [\mathbf{x}_{k,P+1}(P), \dots, \mathbf{x}_{k,P+1}(N-1)], k = 0, 1, \dots, K. \quad (5.19)$$

Similarly,  $\hat{\mathbf{R}}_{yy}(\alpha)$  can be decomposed as

$$\hat{\mathbf{R}}_{yy}(\alpha) = (\mathbf{Y}_0 - \alpha \mathbf{T})(\mathbf{Y}_0 - \alpha \mathbf{T})^H + \sum_{k=1}^K \mathbf{Y}_k \mathbf{Y}_k^H, \quad (5.20)$$

where

$$\mathbf{T} = [\mathbf{t}(P), \dots, \mathbf{t}(N-1)] \in \mathbb{C}^{JP \times (N-P)}, \quad (5.21)$$

$$\mathbf{Y}_k = [\mathbf{y}_k(P), \dots, \mathbf{y}_k(N-1)], k = 0, 1, \dots, K. \quad (5.22)$$

Using (5.17) and (5.20), an asymptotically equivalent expression for (5.14) is derived in Appendix C.1:

$$\begin{aligned} & \ln \left[ 1 + \text{tr} \left\{ (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S})^H \hat{\mathbf{R}}_X^{-1} (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S}) \right\} \right] \\ & - \ln \left[ 1 + \text{tr} \left\{ (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T})^H \hat{\mathbf{R}}_Y^{-1} (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T}) \right\} \right], \end{aligned} \quad (5.23)$$

where

$$\hat{\mathbf{R}}_X = \mathbf{X}_0 \mathbf{P}_S^\perp \mathbf{X}_0^H + \sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^H, \quad (5.24)$$

$$\hat{\mathbf{R}}_Y = \mathbf{Y}_0 \mathbf{P}_T^\perp \mathbf{Y}_0^H + \sum_{k=1}^K \mathbf{Y}_k \mathbf{Y}_k^H, \quad (5.25)$$

with  $\mathbf{P}_S^\perp$  denoting the projection matrix to the orthogonal complement of the range of  $\mathbf{S}^H$

$$\mathbf{P}_S^\perp = \mathbf{I} - \mathbf{P}_S = \mathbf{I} - \mathbf{S}^H (\mathbf{S}^H)^\dagger, \quad (5.26)$$

where  $(\mathbf{S}^H)^\dagger$  denotes the Moore-Penrose pseudo-inverse of  $\mathbf{S}^H$ , while the other projection matrix  $\mathbf{P}_T^\perp$  is similarly defined using the matrix  $\mathbf{T}$ .

Based on the asymptotic expression (5.23), a closed-form estimate of the amplitude is given by (see Appendix C.2).

$$\hat{\alpha} = \frac{\text{tr} \left\{ \mathbf{S}^H \hat{\mathbf{R}}_X^{-1} \mathbf{X}_0 - \mathbf{T}^H \hat{\mathbf{R}}_Y^{-1} \mathbf{Y}_0 \right\}}{\text{tr} \left\{ \mathbf{S}^H \hat{\mathbf{R}}_X^{-1} \mathbf{S} - \mathbf{T}^H \hat{\mathbf{R}}_Y^{-1} \mathbf{T} \right\}}. \quad (5.27)$$

Note that (5.27) is also an asymptotic maximum likelihood (AML) estimate, since the underlying approximations (see Appendices C.1 and C.2) of the likelihood function were made in the asymptotic sense. For convenience, the AML estimator of [7] is henceforth referred to as the **AML1**, whereas the new amplitude estimate (5.27) as the **AML2**. While both estimators are AML, it should be noted that, unlike AML1 that involves a two-step estimation process initialized by the LS estimator, AML2 is in closed-form, requiring only a one-step calculation. As we show in Section 5.5, the two estimators perform similarly when the data size is large; however, in the more challenging case with limited data, AML1 yields a notably worse performance due to the coarse initial estimate provided by the LS estimator.

## 5.4 New Parametric GLRT

Given the AML2 amplitude estimate (5.27), the spatial covariance matrix estimates (7.7) and (7.8) can be obtained in closed-form, which also lead to a closed-form expression of the parametric GLRT test statistic. In particular, we show in Appendix C.3 that the test statistic (7.6) can be expressed as

$$\frac{\left| \sum_{n=P}^{N-1} \left[ \mathbf{s}_{P+1}^H(n) \hat{\mathbf{R}}_X^{-1} \mathbf{x}_{0,P+1}(n) - \mathbf{t}^H(n) \hat{\mathbf{R}}_Y^{-1} \mathbf{y}_0(n) \right] \right|^2}{\sum_{n=P}^{N-1} \left[ \mathbf{s}_{P+1}^H(n) \hat{\mathbf{R}}_X^{-1} \mathbf{s}_{P+1}(n) - \mathbf{t}^H(n) \hat{\mathbf{R}}_Y^{-1} \mathbf{t}(n) \right]} \quad (5.28)$$

=GLR.

To gain additional insight into the test statistic and the behavior of the general parametric GLRT, it is shown in Appendix C.4 that the test statistic can be equivalently expressed as

$$\text{GLR} = \frac{\left| \sum_{n=P}^{N-1} \mathbf{s}_{P+1}^H(n) \mathbf{W} \mathbf{x}_{0,P+1}(n) \right|^2}{\sum_{n=P}^{N-1} \mathbf{s}_{P+1}^H(n) \mathbf{W} \mathbf{s}_{P+1}(n)}, \quad (5.29)$$

where  $\mathbf{s}_{P+1}(n)$  and  $\mathbf{x}_{0,P+1}(n)$  are  $J(P+1) \times 1$  vectors defined in (5.15) and (5.16), and  $\mathbf{W}$  is a block whitening matrix

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 \\ \mathbf{W}_2^H & \mathbf{W}_3 \end{bmatrix}, \quad (5.30)$$

with individual component matrix given by (C.21)-(C.23) of Appendix C.4.

From (5.29), it is seen that the parametric GLRT performs a partial spatio-temporal whitening across  $J(P + 1)$  dimensions (i.e., the size of the regression vectors  $\mathbf{x}_{0,P+1}$  formed from the test signal) using the whitening matrix  $\mathbf{W}$ . Recall that a fully adaptive STAP detector such as Kelly's GLRT [13] performs a joint spatio-temporal whitening across all  $JN$  dimensions, whereas the parametric Rao or PAMF detector performs successive (as opposed to joint) whitening, i.e., temporal whitening followed by spatial whitening [23,40]. Hence, the parametric GLRT is positioned between the two cases. This allows the parametric GLRT to utilize a parametric model and provide data efficiency just like the Rao, meanwhile exploiting more degrees of freedom for more effective interference rejection and detection. This corroborates earlier numerical results [28], which shows that the parametric GLRT in general outperforms the parametric Rao when the data available for estimation becomes very limited.

## 5.5 Numerical Examples

In this section, several simulation results are provided to illustrate the performance of the proposed estimation and detection techniques. We consider simulated data generated using an AR model and the KASSPER data [52] which was obtained from more realistic clutter model. For the first case, the disturbance signal is generated as a multichannel AR(2) process with AR coefficient  $\mathbf{A}$  and spatial covariance matrix  $\mathbf{Q}$ ; these parameters are set to ensure that the AR process is stable and  $\mathbf{Q}$  is a valid covariance matrix, but otherwise randomly selected. The signal vector  $\mathbf{s}$  corresponds to a uniform equispaced linear array with randomly selected normalized spatial and Doppler frequencies. The signal-to-interference plus noise ratio (SINR) is defined as

$$\text{SINR} = |\alpha|^2 \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}. \quad (5.31)$$

### 5.5.1 Estimation

We focus here on the challenging case with zero range training, i.e.,  $K = 0$ , which is of great interest for applications in heterogeneous environments. Under this setup, we consider two sub-cases with: (1)  $N = 32$ , i.e., a moderate value for the number of pulses within a CPI; and (2)  $N = 16$  a more limited scenario. For both cases, we compare the **LS** estimator (5.12), the **AML1** of [28], the **AML2** (5.27), and the **ML** estimator (7.13) initialized by the AML1 (this consideration is motivated by the fact that AML1 provides the best known initial estimate prior to the current work).

Fig. 5.1 presents the mean-squared error (MSE) of the amplitude estimate  $\hat{\alpha}$  obtained by each estimator, along with the Cramér-Rao bound (CRB), a lower bound for any unbiased estimator, versus the SINR when  $N = 32$  and  $J = 4$ . It is seen that, in this case, the AML1 and AML2 amplitude estimates are nearly identical to the ML estimate, while the LS amplitude estimate shows the worse performance among all estimators.

The results for the case of  $N = 16$  is shown in Fig. 5.2. We see that the AML1 estimate is worse than the AML2 estimate in the current case, whereas the ML estimate is the worst due to

inaccurate initialization and local convergence. This clearly shows the limitation of the iterative search based ML estimator.

### 5.5.2 Detection

Here, we report the detection performance under the same setup as in Figs. 5.1 and 5.2. We compare the various parametric GLRT, including **GLRT/AML1** (i.e., the GLRT (7.6) with the AML1 estimator), **GLRT/ML** (the GLRT (7.6) with the ML estimator), and **GLRT/AML2** (the GLRT (5.28) with the new AML2 estimator). Also included in the comparison are the asymptotic result provided by the parametric GLRT (see [28]) and the ideal matched filter (MF) which assumes exact knowledge of  $\mathbf{R}$  and, therefore, cannot be used in practice but offers a baseline for comparison. Here, we set the probability of false alarm  $P_f = 0.01$ .

Fig. 5.3 shows the probability of detection  $P_d$  versus SINR for various detectors when the number of temporal samples  $N = 32$  and no range training data is available. It is seen that the GLRT/AML2 slightly outperforms the GLRT/ML and GLRT/AML1, but overall they are quite similar, and all are within 3 dB from the ideal MF detector. The limited sample case of  $N = 16$  is depicted in Fig. 5.4, where the GLRT/AML2 achieves significantly better results than the GLRT/ML and GLRT/AML1. The poor performance of the GLRT/ML is due to the poor amplitude estimate which, as shown earlier in Fig. 5.2, is caused by inaccurate initialization and local convergence.

### 5.5.3 KASSPER Dataset

In the above simulation, the disturbance is generated by an AR process which matches the assumed model of the parametric detectors. To show the detection performance in a more realistic environment, we use the KASSPER dataset which, first, is not generated from an AR model and, in addition, contains many challenging real-world effects including heterogeneous terrain, array errors, and dense ground targets (see [52] for a detailed description of the KASSPER dataset).

Fig. 5.5 shows the probability of detection versus SINR in the training-free  $K = 0$  case where the number of spatial channels is  $J = 11$  and the number of temporal samples is  $N = 32$ . All parametric detectors use an AR(1) process to model and estimate the disturbance. Results show that the new GLRT/AML2 generally outperforms the GLRT/AML1 and is slightly better than the GLRT/ML at high SINR.

The parametric GLRT effectively trades range training for temporal pulses within a CPI and if the number  $N$  of the latter is large relative to the number of unknowns to be estimated, determined by  $J$  (number of spatial channels) and  $P$  (AR model order). In general, for low order AR models, the parametric GLRT can provide good detection performance (e.g., within 3dB from the MF bound) if  $N/J > 5$  [28]. This is not the case for Fig. 5.5, where  $N/J \approx 3$  and we see a performance gap of about 7 dB. To close the gap, we consider the case when the parametric detectors utilizes  $K = 1$  range training signal while the other parameters are kept the same. There are two guard cells between the test cell and the training cell. In practical radar systems,  $K$  is usually an even number as training data are often taken from both sides of the test cell. Here,

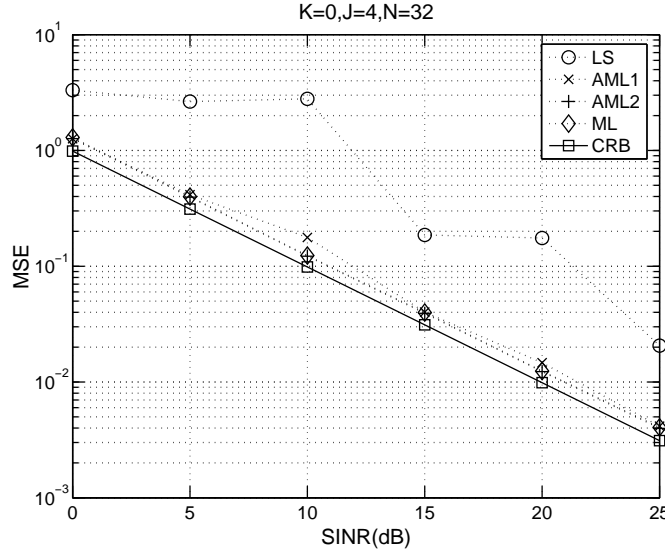


Figure 5.1: MSEs of amplitude estimate  $\hat{\alpha}$  versus SINR when  $J = 4$ ,  $N = 32$ , and  $K = 0$

we choose  $K = 1$  corresponding to a more restrictive case. The results are depicted in Fig. 5.6. It should be noted that in the KASSPER data, clutter across range cells is not i.i.d. [52]. Still, a small amount of training is useful to the parametric detector in the current case, all yielding improved detection performance less than 3 dB from the MF bound. This is due to the fact that the effect of clutter variation across a small area (i.e., for small  $K$ ) is negligible. On the other hand, for a data-demanding non-parametric covariance matrix based STAP detector,  $K$  has to be very large, in which case the effect of range-dependent clutter on such detectors can no longer be neglected [52].

## 5.6 Conclusion

A new parametric GLRT for multichannel adaptive signal detection has been proposed. The detector builds on a new closed-form solution for the underlying nonlinear estimation problem. The new parametric GLRT obviates the need for initial parameter estimation as required by an earlier scheme, is computationally simpler, and provides generally improved detection performance when training data is limited. Due to its data efficiency, our new parametric GLRT and the underlying estimator are particularly useful for detection and estimation in training limited environments

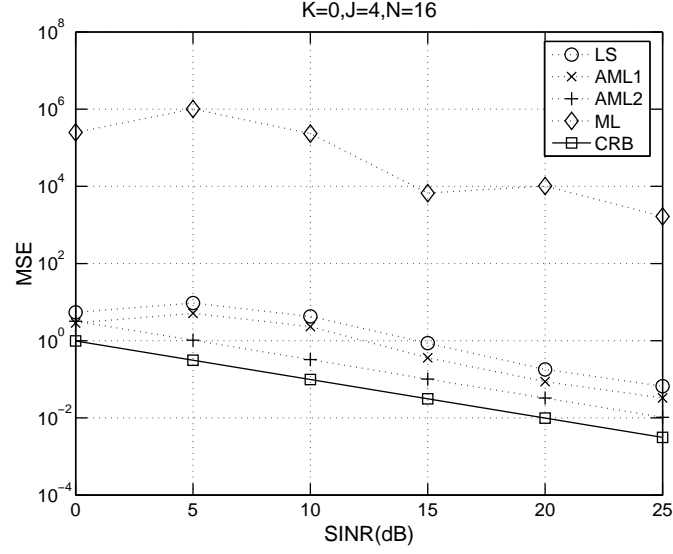


Figure 5.2: MSEs of amplitude estimate  $\hat{\alpha}$  versus SINR when  $J = 4$ ,  $N = 16$ , and  $K = 0$ .

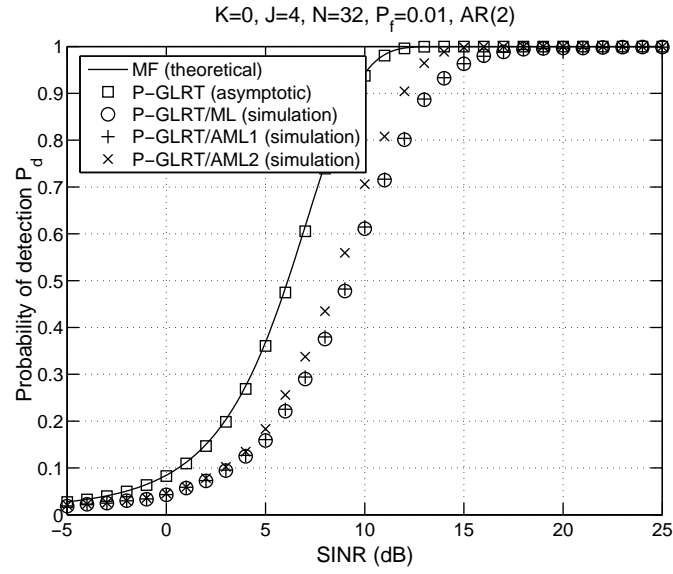


Figure 5.3: Probability of detection  $P_d$  versus SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 32$ , and  $K = 0$ .



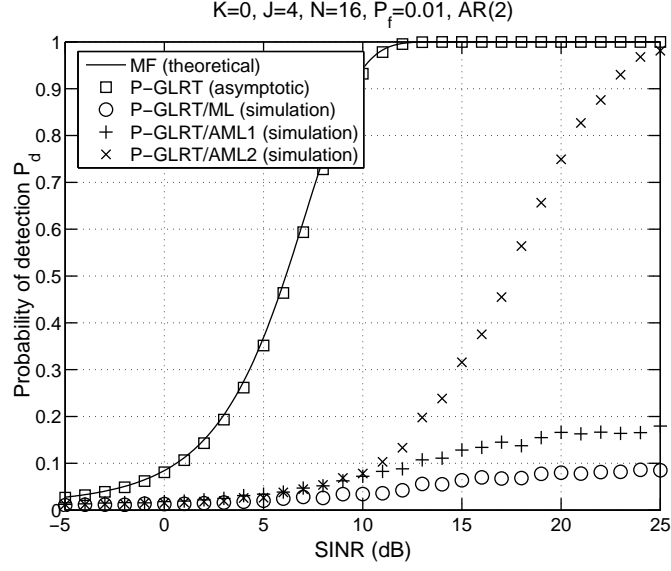


Figure 5.4: Probability of detection  $P_d$  versus SINR when  $P_f = 0.01$ ,  $J = 4$ ,  $N = 16$ , and  $K = 0$ .

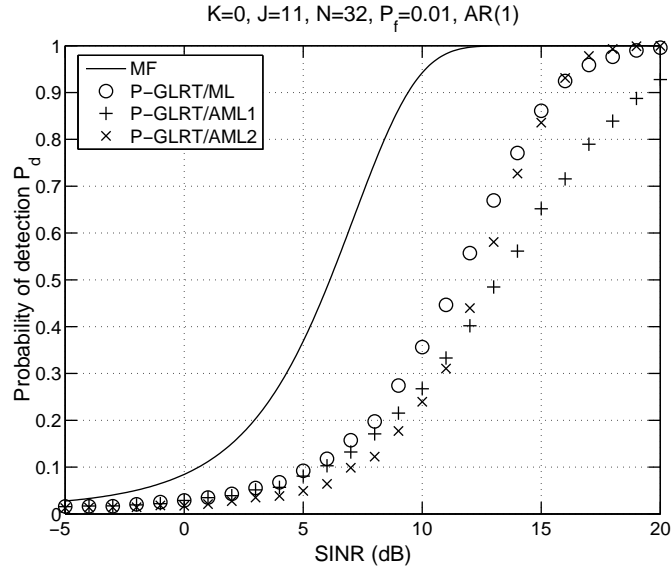


Figure 5.5: Probability of detection  $P_d$  versus SINR for the KASSPER dataset when  $P_f = 0.01$ ,  $J = 11$ ,  $N = 32$ , and  $K = 0$ .

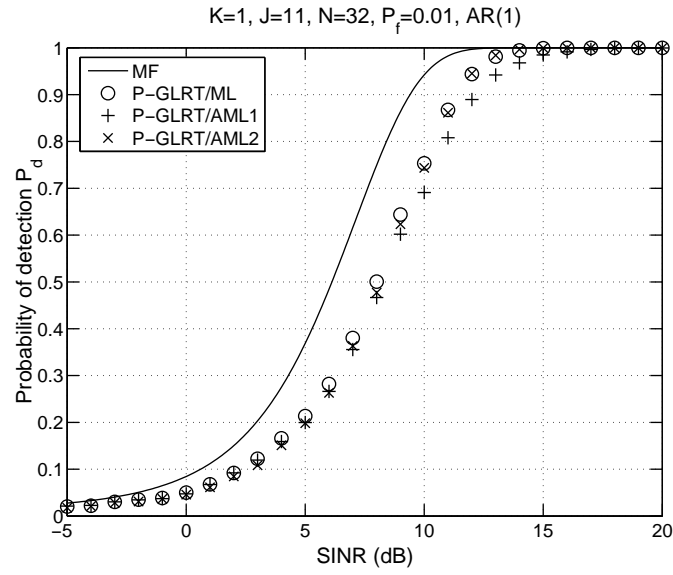


Figure 5.6: Probability of detection  $P_d$  versus SINR for the KASSPER dataset when  $P_f = 0.01$ ,  $J = 11$ ,  $N = 32$ , and  $K = 1$ .

## Chapter 6

# Recursive Parametric Tests for Multichannel Adaptive Signal Detection

### 6.1 Introduction

The parametric Rao and parametric GLRT detectors were developed by assuming that the model order of the multichannel AR process is *known* a priori to the detector. In practice, the model order has to be estimated by some model order selection technique, such as the generalized Akaike Information Criterion (GAIC), Minimum Description Length (MDL), and others [31]. Since most of these model order selection techniques require estimates of the unknown parameters for each possible model order before the best one is identified, a standard non-recursive implementation of the parametric detectors is computationally intensive.

In this chapter, we consider *joint model order selection, parameter estimation, and target detection* for STAP applications. We note that the parameter estimates of a multichannel AR process for all model orders can be efficiently obtained by recursively solving a set of multichannel Yule-Walker equations using the multichannel Levinson algorithm [29, 30]. The multichannel Levinson algorithm yields the parameter estimates for a particular model order at every recursion, following which information criteria such as the GAIC can be conveniently computed. As such, the estimation of the model order is naturally integrated. We follow the above approach and develop recursive versions of the parametric Rao and parametric GLRT detectors. The recursive parametric detectors utilize the Yule-Walker parameter estimates obtained by using the multichannel Levinson algorithm with the biased autocorrelation function (ACF) estimate [29, 30]. Our development of the recursive versions of the parametric detectors integrated with the GAIC for model order selection is well-motivated since the multichannel Levinson algorithm is *computationally efficient* and the model order is *not required to be known* to the detectors. Numerical results show that the Yule-Walker parameter estimates are asymptotically equivalent to the ML estimates originally used in the non-recursive parametric Rao [37] and parametric GLRT [53] detectors. It is also observed that the recursive parametric detectors perform nearly identically to the corresponding non-recursive parametric detectors, even though the formers assume no knowledge of the model order while the latters assume the exact model order.

The rest of the chapter is organized as follows. Section 6.2 contains the data model and problem statement. The non-recursive parametric Rao and parametric GLRT detectors with known model order are summarized in Section 6.3. Section 6.4 contains our recursive parametric Rao and parametric GLRT detectors with unknown model order. Numerical results are presented in Section 6.5, followed by our conclusions in Section 6.6.

## 6.2 Data Model and Problem Statement

Consider the problem of detecting a *known* multichannel signal with *unknown* amplitude in the presence of spatially and temporally colored disturbance (e.g., [1]):

$$\begin{aligned} H_0 : \quad \mathbf{x}_0 &= \mathbf{d}_0, \\ H_1 : \quad \mathbf{x}_0 &= \alpha \mathbf{s} + \mathbf{d}_0, \end{aligned} \tag{6.1}$$

where all vectors are  $JN \times 1$  vectors with  $J$  denoting the number of spatial channels and  $N$  the number of temporal observations. The test signal  $\mathbf{x}_0$  contains a disturbance signal  $\mathbf{d}_0$ , and possibly a target signal  $\alpha \mathbf{s}$ , where  $\mathbf{s}$  denotes the target steering vector which is assumed *known* and  $\alpha$  the *unknown* complex amplitude. In addition to the test signal  $\mathbf{x}_0$ , there may be a set of *target-free range training* or *secondary* signals  $\mathbf{x}_k = \mathbf{d}_k \in \mathbb{C}^{JN \times 1}$ ,  $k = 1, \dots, K$ , that can be exploited to assist in the target detection process. In this chapter, we consider both cases with or without training data; in the latter case, we set  $K = 0$ . The disturbance signals  $\{\mathbf{d}_k\}_{k=0}^K$  are assumed to be independent and identically distributed (i.i.d.) with distribution  $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ , where  $\mathbf{R} \in \mathbb{C}^{JN \times JN}$  is the *unknown* space-time covariance matrix.

Let us decompose the  $JN \times 1$  *space-time* vector  $\mathbf{x}_k$  into a series of  $J \times 1$  *spatial* vectors  $\mathbf{x}_k(n)$  as follows:

$$\mathbf{x}_k = [\mathbf{x}_k^T(0), \dots, \mathbf{x}_k^T(N-1)]^T. \tag{6.2}$$

Let  $\mathbf{d}_k$  and  $\mathbf{s}$  be similarly decomposed into  $\mathbf{d}_k(n) \in \mathbb{C}^{J \times 1}$  and  $\mathbf{s}(n) \in \mathbb{C}^{J \times 1}$ , respectively. Then, we can rewrite the hypothesis testing using the above spatial vectors indexed by  $n$  (time):

$$\begin{aligned} H_0 : \quad \mathbf{x}_0(n) &= \mathbf{d}_0(n), \quad n = 0, \dots, N-1, \\ H_1 : \quad \mathbf{x}_0(n) &= \alpha \mathbf{s}(n) + \mathbf{d}_0(n), \quad n = 0, \dots, N-1. \end{aligned} \tag{6.3}$$

Furthermore, we follow a parametric approach as in [22,23,37,53], which models the disturbance signal  $\mathbf{d}_k(n)$ , as a  $J$ -channel AR( $P$ ) process with *unknown* model order  $P$ :

$$\mathbf{d}_k(n) = - \sum_{i=1}^P \mathbf{A}^H(i) \mathbf{d}_k(n-i) + \boldsymbol{\varepsilon}_k(n), \quad k = 0, 1, \dots, K, \tag{6.4}$$

where  $\{\mathbf{A}^H(i)\}_{i=1}^P$  denote the *unknown*  $J \times J$  AR coefficient matrices and  $\boldsymbol{\varepsilon}_k(n)$  the  $J \times 1$  spatial noise vectors that are temporally white but spatially colored:  $\boldsymbol{\varepsilon}_k(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ , where  $\mathbf{Q} \in \mathbb{C}^{J \times J}$  denotes the *unknown* spatial covariance matrix.

The *problem of interest* is to develop parametric detectors for the above composite hypothesis testing problem (6.1) or (6.3), using the test signal  $\mathbf{x}_0$  and training signals  $\{\mathbf{x}_k\}_{k=1}^K$  (if any). We reiterate that the model order  $P$  is assumed *unknown* to the detector in this chapter whereas the original developments of the PAMF, parametric Rao and parametric GLRT detectors all assume that  $P$  is known [22, 23, 37, 53]. A distinctive feature of this work is that we consider computationally efficient solutions to this joint order selection problem, parameter estimation, and target detection problem.

### 6.3 Non-recursive Parametric Rao and Parametric GLRT Detectors with Known $P$

For easy reference and to facilitate our later development of the recursive parametric detectors, we provide a brief summary of the parametric Rao and GLRT detectors in this section. These detectors are two different solutions to the problem stated in Section 6.2 when the model order  $P$  is *known* [37, 53]. The parametric Rao detector is computationally simpler, but the parametric GLRT offers improved performance. Both detectors first find the ML estimates of the unknown parameters, which are next used to compute the test statistics. The likelihood functions under the null and alternative hypotheses are parameterized by the signal amplitude  $\alpha$ , the AR coefficients  $\mathbf{A}^H = [\mathbf{A}^H(1), \dots, \mathbf{A}^H(P)] \in \mathbb{C}^{J \times JP}$ , and spatial covariance matrix  $\mathbf{Q}$ . Note that under the null hypothesis we have  $\alpha = 0$ . Given  $\mathbf{A}$ , the steering vector and test signal can be temporally whitened through the following inverse (i.e., moving average) filtering:

$$\tilde{\mathbf{s}}(n) = \mathbf{s}(n) + \sum_{i=1}^P \mathbf{A}^H(i) \mathbf{s}(n-i), \quad (6.5)$$

$$\tilde{\mathbf{x}}_0(n) = \mathbf{x}_0(n) + \sum_{i=1}^P \mathbf{A}^H(i) \mathbf{x}_0(n-i). \quad (6.6)$$

This is an important observation exploited by the parametric Rao and parametric GLRT detectors that are summarized next.

The parametric GLRT is given by [53]

$$T_{\text{GLRT}} = 2L \ln \frac{|\hat{\mathbf{Q}}_{\text{ML},0}|}{|\hat{\mathbf{Q}}_{\text{ML},1}|} \underset{H_0}{\overset{H_1}{\geq}} \gamma_{\text{GLRT}}, \quad (6.7)$$

where  $L = (K+1)(N-P)$  and  $\gamma_{\text{GLRT}}$  denotes the corresponding test threshold. The ML estimates of the spatial covariance matrix under the null and alternative hypotheses,  $\hat{\mathbf{Q}}_{\text{ML},0}$  and  $\hat{\mathbf{Q}}_{\text{ML},1}$  are given by

$$\hat{\mathbf{Q}}_{\text{ML},0} = \hat{\mathbf{Q}}(\alpha)|_{\alpha=0}, \quad (6.8)$$

$$\hat{\mathbf{Q}}_{\text{ML},1} = \hat{\mathbf{Q}}(\alpha)|_{\alpha=\hat{\alpha}_{\text{ML}}}. \quad (6.9)$$

The  $\alpha$ -dependent  $\hat{\mathbf{Q}}(\alpha)$  is given by

$$\hat{\mathbf{Q}}(\alpha) = \frac{1}{L} \left( \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha) \right), \quad (6.10)$$

where the  $\alpha$ -dependent correlation matrices are

$$\begin{aligned} \hat{\mathbf{R}}_{xx}(\alpha) &= \sum_{k=1}^K \sum_{n=P}^{N-1} \mathbf{x}_k(n) \mathbf{x}_k^H(n) \\ &\quad + \sum_{n=P}^{N-1} [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)] [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)]^H, \end{aligned} \quad (6.11)$$

$$\begin{aligned} \hat{\mathbf{R}}_{yy}(\alpha) &= \sum_{k=1}^K \sum_{n=P}^{N-1} \mathbf{y}_k(n) \mathbf{y}_k^H(n) \\ &\quad + \sum_{n=P}^{N-1} [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)] [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)]^H, \end{aligned} \quad (6.12)$$

$$\begin{aligned} \hat{\mathbf{R}}_{yx}(\alpha) &= \sum_{k=1}^K \sum_{n=P}^{N-1} \mathbf{y}_k(n) \mathbf{x}_k^H(n) \\ &\quad + \sum_{n=P}^{N-1} [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)] [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)]^H, \end{aligned} \quad (6.13)$$

with  $\mathbf{t}(n)$  and  $\mathbf{y}(n)$  denoting the regression subvectors formed from the steering vector  $\mathbf{s}(n)$  and test signal  $\mathbf{x}_0(n)$ , respectively:  $\mathbf{t}(n) = [\mathbf{s}^T(n-1), \dots, \mathbf{s}^T(n-P)]^T \in \mathbb{C}^{JP \times 1}$  and  $\mathbf{y}(n) = [\mathbf{x}_0^T(n-1), \dots, \mathbf{x}_0^T(n-P)]^T \in \mathbb{C}^{JP \times 1}$ . The ML estimate of  $\alpha$  under the alternative hypothesis, which is used in (6.9), is given by

$$\hat{\alpha}_{\text{ML}} = \arg \min_{\alpha} \left| \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha) \right|. \quad (6.14)$$

The parametric Rao test is given by [37]

$$T_{\text{Rao}} = \frac{2 \left| \sum_{n=P}^{N-1} \hat{\mathbf{s}}^H(n) \hat{\mathbf{Q}}_{\text{ML},0}^{-1} \hat{\mathbf{x}}_0^H(n) \right|^2}{\sum_{n=P}^{N-1} \hat{\mathbf{s}}^H(n) \hat{\mathbf{Q}}_{\text{ML},0}^{-1} \hat{\mathbf{s}}^H(n)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{Rao}}, \quad (6.15)$$

where  $\gamma_{\text{Rao}}$  denotes the test threshold. The temporally whitened steering vector  $\hat{\mathbf{s}}^H(n)$  and test signal  $\hat{\mathbf{x}}_0^H(n)$  are obtained by replacing  $\mathbf{A}^H$  with the ML estimates under  $H_0$

$$\hat{\mathbf{A}}_{\text{ML},0}^H = -\hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha)|_{\alpha=0}, \quad (6.16)$$

in (6.5) and (6.6), respectively.

The Rao test is asymptotically equivalent to the GLRT but may be inferior to the latter when the data size is small. In addition, the Rao test is obtained based on a low order Taylor expansion of the GLRT, an approximation which is only valid for weak signals [7]. As such, the performance of the parametric Rao detector degrades when the weak signal assumption is violated. The parametric GLRT was developed as an improved detector to deal with the above issues. However, the cost function of the ML amplitude estimator in (6.14) is highly nonlinear. Newton-like iterative nonlinear searches are generally used to find the ML amplitude estimate. Another suboptimum but computationally more efficient estimator, referred to as the asymptotic ML (AML) estimator, was developed in [53]. The AML estimator, which was found to yield similar performance to the ML estimator, can be implemented as follows:

- **Step 1** First, compute a least-squares (LS) amplitude estimate  $\hat{\alpha}_{\text{LS}} = \frac{\mathbf{s}^H \mathbf{x}_0}{\mathbf{s}^H \mathbf{s}}$ . Then, determine an estimate  $\hat{\mathbf{A}}_{\text{LS}}^H$  of  $\mathbf{A}^H$  as follows:

$$\hat{\mathbf{A}}_{\text{LS}}^H = \hat{\mathbf{R}}_{yx}^H(\hat{\alpha}_{\text{LS}}) \hat{\mathbf{R}}_{yy}^{-1}(\hat{\alpha}_{\text{LS}}), \quad (6.17)$$

which can be shown to be statistically consistent [53].

- **Step 2** Compute the temporally whitened signals  $\hat{\tilde{\mathbf{x}}}_k(n)$  and  $\hat{\tilde{\mathbf{s}}}(n)$  by replacing  $\mathbf{A}^H$  with the LS AR coefficient estimate  $\hat{\mathbf{A}}_{\text{LS}}^H$  in (6.5) and (6.6), respectively. Then, obtain the AML amplitude estimate  $\hat{\alpha}_{\text{AML}}$  by using

$$\hat{\alpha}_{\text{AML}} = \frac{\text{tr} \left( \hat{\tilde{\mathbf{S}}}^H \boldsymbol{\Psi}^{-1} \hat{\tilde{\mathbf{X}}}_0 \right)}{\text{tr} \left( \hat{\tilde{\mathbf{S}}}^H \boldsymbol{\Psi}^{-1} \hat{\tilde{\mathbf{S}}} \right)}, \quad (6.18)$$

where  $\hat{\tilde{\mathbf{S}}} = [\hat{\tilde{\mathbf{s}}}(P), \dots, \hat{\tilde{\mathbf{s}}}(N-1)] \in \mathbb{C}^{J \times (N-P)}$ ,  $\hat{\tilde{\mathbf{X}}}_k = [\hat{\tilde{\mathbf{x}}}_k(P), \dots, \hat{\tilde{\mathbf{x}}}_k(N-1)] \in \mathbb{C}^{J \times (N-P)}$ , and

$$\boldsymbol{\Psi} = \hat{\tilde{\mathbf{X}}}_0 \mathbf{P}^\perp \hat{\tilde{\mathbf{X}}}_0^H + \sum_{k=1}^K \hat{\tilde{\mathbf{X}}}_k \hat{\tilde{\mathbf{X}}}_k^H, \quad (6.19)$$

with  $\mathbf{P}^\perp$  denoting the projection matrix projecting to the orthogonal complement of the range of  $\hat{\tilde{\mathbf{S}}}^H$ :  $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P} = \mathbf{I} - \hat{\tilde{\mathbf{S}}}^H \left( \hat{\tilde{\mathbf{S}}}^H \right)^\dagger \in \mathbb{C}^{(N-P) \times (N-P)}$ .

- **Step 3** Find the AML estimate of the spatial covariance matrix by substituting  $\hat{\alpha}_{\text{AML}}$  for  $\alpha$  in (6.10).

Recall that the parametric Rao and parametric GLRT detectors utilize both the test and training signals for the parameter estimation. As a result, they are functional even without training data [53]. The capability to handle the training-free detection is a unique and desirable attribute of the parametric detectors which is not shared by other existing detectors including the PAMF detector. Nevertheless, we need a way to efficiently find an accurate estimate of the model order  $P$ .

## 6.4 Recursive Parametric Tests with Unknown $P$

A standard non-recursive implementation of the parametric detectors is computationally intensive since the parameter estimation for the underlying parametric model has to be repeated for all possible model orders before the best one is identified. Therefore, there is a need to develop more efficient solutions for joint model order selection, parameter estimation, and detection.

We present herein recursive versions of the parametric Rao and parametric GLRT detectors. The multichannel Levinson algorithm is used to recursively solve a set of multichannel Yule-Walker equations for model order  $p = 1, 2, \dots, P_{\max}$ , where  $P_{\max}$  is an upper bound on the model order  $P$ . Interestingly, the complexity involved in the above procedure, which provides parameter estimates for all  $P_{\max}$  model orders, has lower complexity than that involved in solving a single model order  $p = P_{\max}$  by the ML approach (see Section 6.4.5 for details). Given these parameter estimates for all possible  $p$ , an information criterion such as the GAIC can be conveniently utilized to identify the best model order as well as the associated estimates of  $\mathbf{A}$ ,  $\mathbf{Q}$ , and  $\alpha$ . These parameter estimates are then used to compute the final test statistics for the parametric Rao and GLRT detectors. In the following, we discuss the details of the proposed joint approach.

### 6.4.1 Parameter Estimation by the Multichannel Levinson Algorithm

Assume that signal  $\mathbf{x}(n)$  is a  $J$ -channel AR( $P$ ) process as described in (6.4). Estimates of the unknown parameters can be obtained by solving the multichannel Yule-Walker equations given by [29, 30]

$$\mathbf{R}(m) = \begin{cases} -\sum_{i=1}^P \mathbf{A}^H(i) \mathbf{R}(m-i), & m \geq 1, \\ -\sum_{i=1}^P \mathbf{A}^H(i) \mathbf{R}(-i) + \mathbf{Q}, & m = 0, \end{cases} \quad (6.20)$$

where the autocorrelation matrix is defined as

$$\mathbf{R}(m) = E[\mathbf{x}(n) \mathbf{x}^H(n-m)]. \quad (6.21)$$

In matrix form, the multichannel Yule-Walker equations become

$$\mathcal{A}_P \mathcal{R} = [\mathbf{Q} \ 0 \ \dots \ 0], \quad (6.22)$$

where the block matrix  $\mathcal{A}_P$  contain the multichannel AR coefficients and  $\mathcal{R}$  is a block Toeplitz matrix:

$$\mathcal{A}_P = [\mathbf{I} \ \mathbf{A}_P^H(1) \ \dots \ \mathbf{A}_P^H(P)], \quad (6.23)$$

$$\mathcal{R} = \begin{bmatrix} \mathbf{R}(0) & \dots & \mathbf{R}(P) \\ \mathbf{R}(-1) & \dots & \mathbf{R}(P-1) \\ \vdots & \ddots & \vdots \\ \mathbf{R}(-P) & \dots & \mathbf{R}(0) \end{bmatrix}. \quad (6.24)$$



The multichannel Levinson algorithm can be used to recursively solve the above multichannel Yule-Walker equations for different model orders as follows [29, 30].

The multichannel Levinson algorithm begins with the following initial conditions:

$$\mathbf{Q}_0^f = \mathbf{Q}_0^b = \mathbf{R}(0), \quad (6.25)$$

$$\mathcal{A}_0 = \mathcal{B}_0 = \mathbf{I}. \quad (6.26)$$

Henceforth, superscripts  $f$  and  $b$  denote the forward and backward directions of a linear prediction process used by the Levinson algorithm, subscript denotes the order of the linear predictor, and  $\mathcal{A}$  and  $\mathcal{B}$  denote the block row matrices formed by the forward and backward AR coefficient matrices, respectively.

Given the  $p$ -th order forward and backward AR coefficient matrices  $\mathcal{A}_p$  and  $\mathcal{B}_p$ , the forward and backward reflection coefficient matrices for the  $(p+1)$ -st order linear predictors are computed by

$$\mathbf{K}_{p+1}^{fH}(p+1) = -\Delta_{p+1}(\mathbf{Q}_p^b)^{-1}, \quad (6.27)$$

$$\mathbf{K}_{p+1}^{bH}(p+1) = -\nabla_{p+1}(\mathbf{Q}_p^f)^{-1}, \quad (6.28)$$

where  $\Delta_{p+1}$  and  $\nabla_{p+1}$  are defined as

$$\Delta_{p+1} = \sum_{i=0}^p \mathbf{K}_p^{fH}(i) \mathbf{R}(p+1-i), \quad (6.29)$$

$$\nabla_{p+1} = \sum_{i=0}^p \mathbf{K}_p^{bH}(i) \mathbf{R}(i-p-1). \quad (6.30)$$

Next, we update the forward and backward AR coefficient matrices for the  $(p+1)$ -st order predictors as follows:

$$\mathcal{A}_{p+1} = [\mathcal{A}_p, \mathbf{0}] + \mathbf{K}_{p+1}^{fH}(p+1) [\mathbf{0}, \mathcal{B}_p], \quad (6.31)$$

$$\mathcal{B}_{p+1} = [\mathbf{0}, \mathcal{B}_p] + \mathbf{K}_{p+1}^{bH}(p+1) [\mathcal{A}_p, \mathbf{0}]. \quad (6.32)$$

Finally, we update the forward and backward prediction error covariance matrices for the  $(p+1)$ -st order predictors:

$$\mathbf{Q}_{p+1}^f = \mathbf{Q}_p^f + \mathbf{K}_{p+1}^{fH}(p+1) \nabla_{p+1}, \quad (6.33)$$

$$\mathbf{Q}_{p+1}^b = \mathbf{Q}_p^b + \mathbf{K}_{p+1}^{bH}(p+1) \Delta_{p+1}, \quad (6.34)$$

which completes the  $p$ -th recursion of the multichannel Levinson algorithm. Note that the solutions to the  $p$ -th order multichannel Yule-Walker equations are  $\mathcal{A}_p$ , and  $\mathbf{Q}_p^f$ .

In practice, the space-time covariance matrix  $\mathbf{R}(m)$  in the multichannel Yule-Walker equations should be replaced by some estimate. The biased ACF estimate given by

$$\hat{\mathbf{R}}(m) = \frac{1}{N} \sum_{n=0}^{N-1-m} \mathbf{x}(n+m) \mathbf{x}^H(n), \quad (6.35)$$

is usually recommended since it guarantees that the  $\mathcal{R}$  is nonnegative definite [29, 30].

### 6.4.2 AR Model Order Selection

Model order selection for parametric models is a classical research topic and has been investigated for various models (e.g., [30, 31], and references therein). Herein, we consider the GAIC, which has been observed to yield good performance for model order selection (e.g., [54]). The GAIC chooses the model order  $p$  that minimizes

$$W(p) = V(p) + \eta(p), \quad (6.36)$$

where  $V(p)$  is the minimum negative log likelihood function and  $\eta(p)$  is a penalty term that penalizes increasing model order [31]. The minimum negative log likelihood function can be shown to be

$$V(p) = J(K+1)(N-p) \ln(e\pi) + (K+1)(N-p) \ln |\hat{\mathbf{Q}}|, \quad (6.37)$$

where the dependence on  $p$  is made explicit. The penalty term typically takes the form as [31]

$$\eta(p) = 2cJ^2p \ln(\ln(K+1)(N-p)), \quad (6.38)$$

where  $c \geq 1$  is a parameter of user choice. It has been found that (6.36) along with (6.38) usually provides a consistent model order estimation [31].

### 6.4.3 Recursive Parametric Rao Test

Based on the above recursive parameter estimation and model order selection techniques, the parametric Rao test can be implemented in a recursive manner as follows:

- **Step 1** Obtain the biased ACF estimate according to (6.35):

$$\hat{\mathbf{R}}(m) = \frac{1}{N(K+1)} \sum_{k=0}^K \sum_{n=0}^{N-1-m} \mathbf{x}_k(n+m) \mathbf{x}_k^H(n), \quad (6.39)$$

$$m = 0, 1, \dots, P_{\max}.$$

Note that both the training and test signals are used to obtain the ACF estimate.

- **Step 2** Initialization: Set  $p = 0$  and initialize the forward and backward prediction error covariance matrices,  $\mathbf{Q}_0^f$  and  $\mathbf{Q}_0^b$ , and the forward and backward AR coefficient matrices,  $\mathcal{A}_0$  and  $\mathcal{B}_0$ , as in (6.25) and (6.26). Compute the GAIC  $W(0)$  for the 0-th model order by using (6.36).
- **Step 3a** Compute the forward and backward reflection coefficient matrices for the  $(p+1)$ -st order linear predictors,  $\mathbf{K}_{p+1}^{fH}(p+1)$  and  $\mathbf{K}_{p+1}^{bH}(p+1)$ , by using (6.27) and (6.28).
- **Step 3b** Update the forward and backward AR coefficient matrices for the  $(p+1)$ -st order predictors,  $\mathcal{A}_{p+1}$  and  $\mathcal{B}_{p+1}$ , by using (6.31) and (6.32). Update the forward and backward prediction error covariance matrices for the  $(p+1)$ -st order predictors,  $\mathbf{Q}_{p+1}^f$  and  $\mathbf{Q}_{p+1}^b$ , by using (6.33) and (6.34).

- **Step 3c** Compute the GAIC  $W(p+1)$  for the  $(p+1)$ -st model order, by using (6.36).
  - If  $p = 0$ , increase  $p$  by 1 and go back to **Step 3a**;
  - else if  $W(p+1) \geq W(p)$ , go to **Step 4**;
  - otherwise, increase  $p$  by 1 and go back to **Step 3a**.

The following upper bound can be imposed for model order selection [23]

$$p \leq \left\lfloor \frac{3\sqrt{N}}{J} \right\rfloor. \quad (6.40)$$

- **Step 4** The order estimate  $\hat{P}$  is  $p$  (i.e., the final value of the above recursion index). For the selected model order  $\hat{P} = p$ , obtain the parameter estimates:

$$\hat{\mathbf{A}}^H(i) = \mathbf{A}_{\hat{P}}^H(i), \quad i = 1, 2, \dots, \hat{P}, \quad (6.41)$$

$$\hat{\mathbf{Q}} = \mathbf{Q}_{\hat{P}}^f. \quad (6.42)$$

Compute the parametric Rao test statistic (6.15) by replacing the ML parameter estimates (6.16) and (6.8) with the obtained Yule-Walker solutions (6.41) and (6.42), respectively. Finally, the test statistic is compared with a test threshold to decide if the target is present. The test threshold can be determined by using the asymptotic analysis in [37].

#### 6.4.4 Recursive Parametric GLRT

Recursive implementation of the parametric GLRT is more involved than the recursive parametric Rao test. The reason is that finding the ML estimate of signal amplitude  $\alpha$ , which is required by the parametric GLRT, is nonlinear even with a known model order [53, 55]. To circumvent the problem, we consider a recursive parametric GLRT by using the model order estimate obtained by the recursive parametric Rao test.

The recursive implementation of the parametric GLRT can be summarized as follows:

- **Step 1** Find the spatial covariance matrix estimate  $\hat{\mathbf{Q}}_0$  under  $H_0$  and the model order estimate  $\hat{P}$  by using the multichannel Levinson algorithm in the same manner as in the recursive parametric Rao test.
- **Step 2** Using the model order estimate  $\hat{P}$  obtained in **Step 1**, find the amplitude estimate  $\hat{\alpha}$  by either (6.14) or (6.18). Next, obtain the spatial covariance matrix estimate  $\hat{\mathbf{Q}}_1$  by using  $\hat{\alpha}$  and  $\hat{P}$ . Specifically, the spatial covariance matrix estimate  $\hat{\mathbf{Q}}_1$  can be obtained by running the multichannel Levinson algorithm a second time (with  $\hat{P}$  recursions) along with the following modified ACF estimate:

$$\begin{aligned} \hat{\mathbf{R}}(m) = \frac{1}{N(K+1)} & \left\{ \sum_{n=0}^{N-1-m} \check{\mathbf{x}}_0(n+m) \check{\mathbf{x}}_0^H(n) \right. \\ & \left. + \sum_{k=1}^K \sum_{n=0}^{N-1-m} \mathbf{x}_k(n+m) \mathbf{x}_k^H(n) \right\}, \end{aligned} \quad (6.43)$$

where  $\check{\mathbf{x}}_0(n) = \mathbf{x}_0(n) - \hat{\alpha}\mathbf{s}(n)$ .

- **Step 3** Compute the test statistic (6.7) by replacing the ML parameter estimates (6.8) and (6.9) with the Yule-Walker solutions  $\hat{\mathbf{Q}}_0$  and  $\hat{\mathbf{Q}}_1$ , respectively. Finally, the test statistic is compared with a test threshold to decide if the target is present.

### 6.4.5 Complexity

We provide a brief discussion on the complexity involved in the recursive parametric Rao test versus its non-recursive counterpart. Since the recursive and non-recursive implementations differ only in parameter estimation (they share identical steps in signal whitening and calculating the test statistic), we only compare the complexity involved in finding estimates of the AR coefficients  $\mathbf{A}$  and the spatial covariance matrix  $\mathbf{Q}$ . Tables 6.1 and 6.2 contain a summary of the number of flops involved in the major steps of the recursive and, respectively, non-recursive parameter estimation. In general, we have  $(K + 1)N > JP_{\max}$  for practical applications. Then, it can be concluded from Tables 6.1 and 6.2 that the recursive parameter estimation, which yields parameter estimates for all model orders, has a overall complexity of  $O(J^2P_{\max}(K + 1)N)$ , whereas the overall complexity of the non-recursive estimation for is  $O(J^2P_{\max}^3(K + 1)N)$ , which is  $P_{\max}^2$  times higher.

Similar conclusions can be made for the parametric GLRT since, just like the parametric Rao test, the recursive and non-recursive implementations differ only in how parameter estimates are obtained.

## 6.5 Numerical Results

In this section, we present simulation results to illustrate the performance of the proposed techniques. The disturbance signal is generated as a multichannel AR(2) process (i.e.,  $P = 2$ ) with randomly selected AR coefficients  $\mathbf{A}$  and a spatial covariance matrix  $\mathbf{Q}$ . These parameters are set to ensure that the AR process is stable and  $\mathbf{Q}$  is a valid covariance matrix, but otherwise randomly selected. The steering vector  $\mathbf{s}$  corresponds to a uniform equi-spaced linear array with  $J = 4$  and randomly selected normalized spatial and Doppler frequencies (see [23]). The signal-to-interference-plus-noise ration (SINR) is defined as

$$\text{SINR} = |\alpha|^2 \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}, \quad (6.44)$$

where the  $JN \times JN$  space-time covariance matrix can be uniquely determined once  $\mathbf{A}^H$  and  $\mathbf{Q}$  are selected.

### 6.5.1 Estimation

We first examine the estimation performance of the solutions to the multichannel Yule-Walker equations. Since  $\mathbf{Q}$  is a matrix, we define the following metric:

$$\xi(\mathbf{Q}) = \frac{1}{J^2} \text{tr} \left\{ E \left[ \left( \hat{\mathbf{Q}} - \mathbf{Q} \right)^H \left( \hat{\mathbf{Q}} - \mathbf{Q} \right) \right] \right\}, \quad (6.45)$$

which is the average of the mean squared errors (MSEs) of all elements of the matrix. For brevity, the above metric is referred to as the MSEs henceforth.

Figures 6.1 and 6.2 depict the MSE of the spatial covariance matrix estimate  $\mathbf{Q}$  versus the number of temporal observations  $N$ . We consider Yule-Walker estimate obtained by using the multichannel Levinson algorithm with the corresponding ML estimate (6.8). Figure 6.1 shows the case without training data ( $K = 0$ ), while figure 6.2 corresponds to the case with limited training data ( $K = 2$ ). It is observed that the Yule-Walker estimate is asymptotically (for large  $N$  and/or large  $K$ ) equivalent to the ML estimate, while the performance of the Yule-Walker estimate may be different when the data size is small. Figure 6.1 shows that the Yule-Walker estimate performs slightly better than the ML estimate, when the number of the temporal observations  $N$  is small and no training data are available ( $K = 0$ ). Although it is generally believed that the ML estimate is more accurate than the Yule-Walker estimate (e.g., [56]), with an extremely small data size as considered in this example (e.g.,  $N = 10, K = 0$ ), either one of the two estimators can slightly outperform the other depending on the choice of the AR parameters. It should also be noted that the bias of the Yule-Walker estimate (because of the use of the biased ACF estimate) can be significant [56]. Figure 6.2 shows that the Yule-Walker estimate performs nearly identically to the ML estimate when training data are used ( $K = 2$ ). It is also observed that the Yule-Walker and ML estimates improve as the training data ( $K$ ) and/or temporal observations ( $N$ ) increases, i.e., the data size increases.

### 6.5.2 Detection

We next examine the detection performance of the recursive parametric Rao and GLRT detectors. For the recursive parametric GLRT detector, the AML, instead of the ML, estimate of  $\alpha$  is used since the detection difference between the two is negligible (see [53]) while the AML is computationally simpler. Also included in comparison is the ideal matched filter (MF), which assumes the exact knowledge of  $\mathbf{R}$  and, therefore, cannot be used in practice but offers a baseline for comparison. In all examples, we set the probability of false alarm  $P_f = 0.01$ . Recursive and Non-Recursive are denoted by ‘R’ and ‘NR’, respectively, in the figures. For example, recursive and non-recursive parametric GLRT detectors are denoted by R-GLRT and NR-GLRT, respectively.

Figures 6.3 to 6.6 depict the probability of detection of various detectors versus the SINR for the recursive parametric detector with unknown model order  $P$  and their non-recursive counterparts with known  $P$ . Figures 6.3 and 6.4 show the case without training data ( $K = 0$ ), and Figures 6.5 and 6.6 correspond to the case with limited training data ( $K = 2$  and 8). We see

Equation	Flops	Remark
(6.39)	$O(J^2 P_{\max}(K+1)N)$	one time calculation
(6.27), (6.28)	$O(J^3(p+2))$	at $p$ th recursion
(6.31), (6.32)	$O(J^3 p)$	at $p$ th recursion
(6.33), (6.34)	$O(J^3)$	at $p$ th recursion
Subtotal	$O(J^3 p)$	at $p$ th recursion
Total	$O(J^2 P_{\max}(K+1)N) + O(J^3 P_{\max}^2)$ $\approx O(J^2 P_{\max}(K+1)N)$	for $p = 1, \dots, P_{\max}$

Table 6.1: Complexity of the Yule-Walker estimator with the multichannel Levinson algorithm for model orders  $p = 1, \dots, P_{\max}$  (recursive implementation)

Equation	Flops	Remark
(6.11)	$O(J^2(K+1)(N-p))$	for model order $p$
(6.12)	$O(J^2 p^2(K+1)(N-p))$	for model order $p$
(6.13)	$O(J^2 p(K+1)(N-p))$	for model order $p$
(6.10)	$O(J^3(p^3 + p^2 + p))$	for model order $p$
Subtotal	$O(J^2 p^2(K+1)N) + O(J^3 p^3)$ $\approx O(J^2 p^2(K+1)N)$	for model order $p$
Total	$O(J^2 P_{\max}^3(K+1)N)$	for $p = 1, \dots, P_{\max}$

Table 6.2: Complexity of the ML estimator for model orders  $p = 1, \dots, P_{\max}$  (non-recursive implementation)

that in general, the performance of the recursive parametric detectors with unknown  $P$  is nearly identical to that of their non-recursive counterpart with known  $P$ . This is particular true for the cases shown in Figures 6.4 to 6.6, where the data size is relatively large (large  $N$  with  $K = 0$ , or a moderate  $N$  with non-zero  $K$ ). This is probably because the Yule-Walker parameter estimate is slightly more accurate than the ML estimate in this case (see Figure 6.1).

## 6.6 Conclusions

We have presented recursive versions of the parametric Rao and parametric GLRT detectors, utilizing the multichannel Levinson algorithm to solve the multichannel Yule-Walker equations recursively and find the estimates of the unknown parameters, along with a GAIC for model order selection. Numerical results show that the Yule-Walker estimate obtained by using the multichannel Levinson algorithm along with the biased ACF estimate is asymptotically equivalent to the ML estimate originally used in the non-recursive parametric Rao and parametric GLRT detectors. It is also shown that the proposed recursive parametric detectors that assume no knowledge about the model order perform nearly identically to the corresponding non-recursive parametric detectors with perfect knowledge of the model order, while the formers have reduced computational complexity.

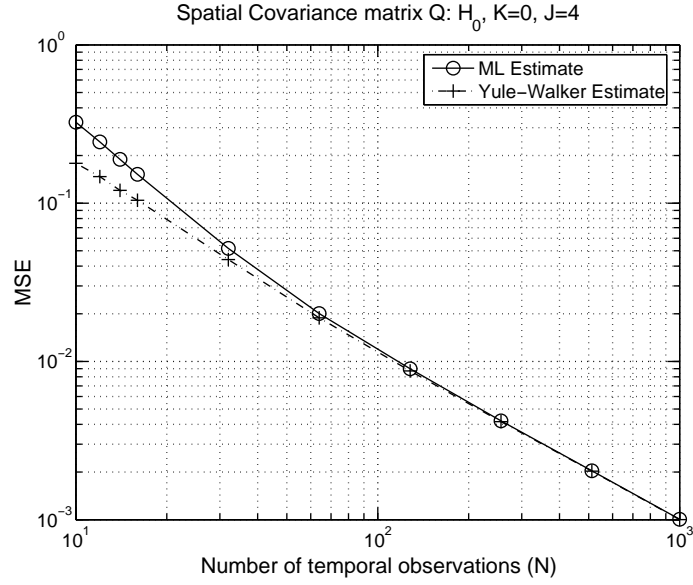


Figure 6.1: MSE of spatial covariance matrix estimate versus the number of temporal observations  $N$  when  $K = 0$  and  $J = 4$ .

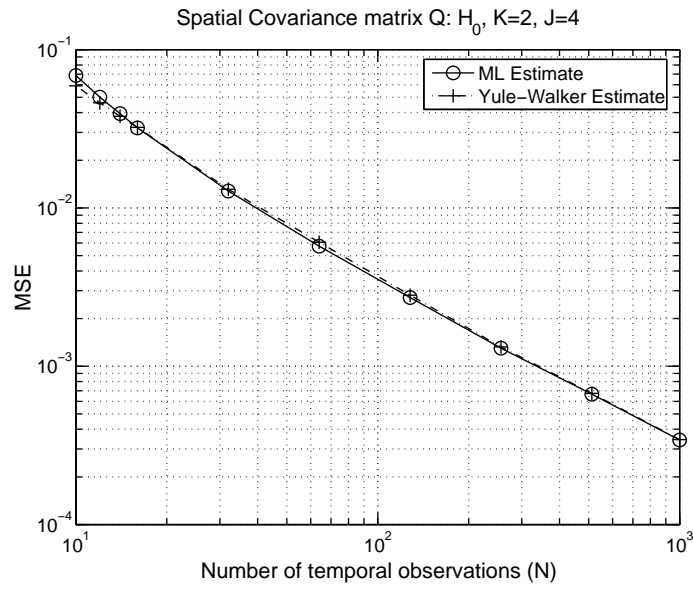


Figure 6.2: MSE of spatial covariance matrix estimate versus the number of temporal observations  $N$  when  $K = 2$  and  $J = 4$ .

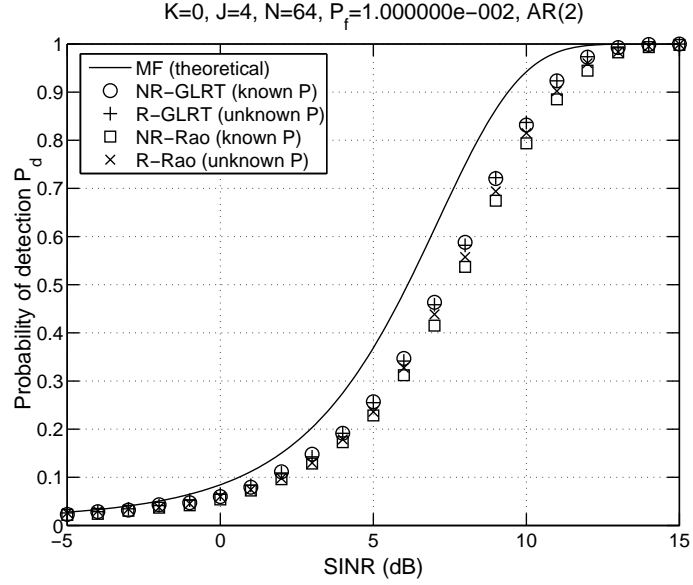


Figure 6.3: Probability of detection  $P_d$  versus SINR when  $K = 0$ ,  $J = 4$ ,  $N = 64$ ,  $P = 2$ , and  $P_f = 0.01$ .

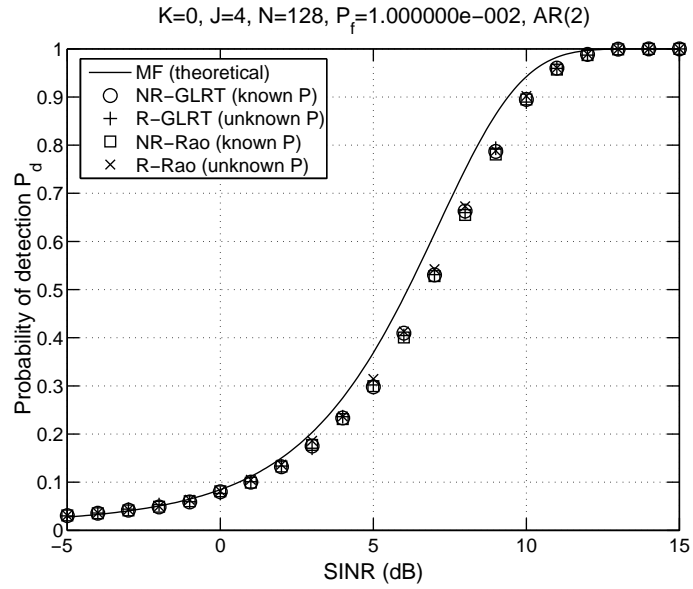


Figure 6.4: Probability of detection  $P_d$  versus SINR when  $K = 0$ ,  $J = 4$ ,  $N = 128$ ,  $P = 2$ , and  $P_f = 0.01$ .



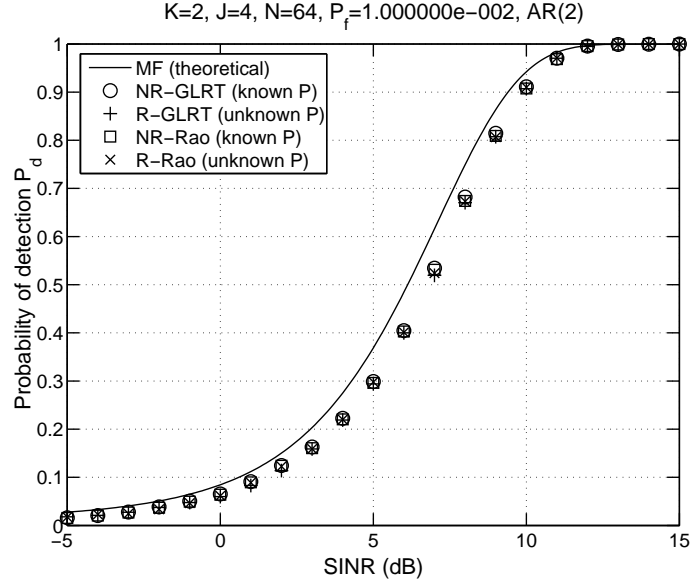


Figure 6.5: Probability of detection  $P_d$  versus SINR when  $K = 2$ ,  $J = 4$ ,  $N = 64$ ,  $P = 2$ , and  $P_f = 0.01$ .

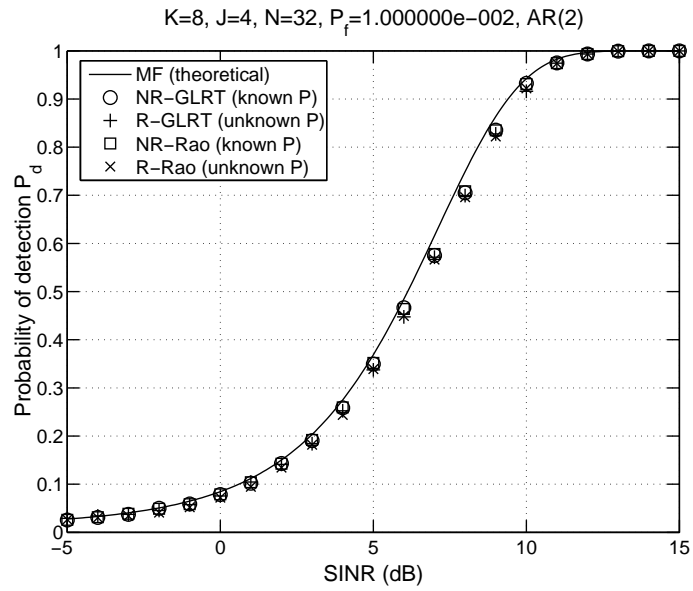


Figure 6.6: Probability of detection  $P_d$  versus SINR when  $K = 8$ ,  $J = 4$ ,  $N = 32$ ,  $P = 2$ , and  $P_f = 0.01$ .

## Chapter 7

# Performance Evaluation of Parametric Space-Time Adaptive Detectors

### 7.1 Introduction

Parametric model based STAP detectors have attracted significant interest in addressing the problem of limited training [22–24, 28, 40]. Specifically, the *parametric adaptive matched filter* (PAMF) [23], which was developed by exploiting a multichannel AR model for the disturbance, significantly outperforms the aforementioned conventional STAP detectors at a reduced complexity when the training size is small. Recently, the PAMF detector has been shown to be equivalent to a *parametric Rao detector* with one exception: while the original PAMF uses only training signals for parameter estimation, the parametric Rao detector uses both training and test signals for estimation [40]. Moreover, another parametric detector, referred to as the *parametric GLRT*, offers improved detection performance over the parametric Rao test when data is limited [28]. Computer simulations show that the parametric Rao and GLRT detectors work well with *limited or even no range training data*. The parametric Rao and GLRT detectors utilize both test and training data for parameter estimation; in the absence of training, the parameters are estimated solely from the test signal [28]. The capability to handle the training-free detection problem is a unique and desirable attribute of these parametric STAP detectors, making them good candidates for detection in heterogeneous environments.

To facilitate the development of the parametric Rao and GLRT detectors, two assumptions were made, which may not hold exactly in real airborne radar systems [19,21]. The first assumption is that the disturbance follows a multichannel AR process. The second assumption is that the training signals, if available, are assumed to be independent and identically distributed (i.i.d.). In practice, the training data is subject to contamination by clutter discretely and/or interfering targets, in which case the training data becomes heterogeneous. It is noted that the performance of the parametric Rao and GLRT detectors was evaluated through computer simulations [28,40] where the disturbance signals were generated to meet the aforementioned assumptions. Therefore, it would be interesting to find out how these parametric detectors perform in real radar environments and provide insight to its ability to real application in airborne radars.

In this chapter, we examine the detection performance of the parametric Rao and GLRT detectors using three more realistic datasets: 1) the Knowledge-Aided Sensor Signal Processing and Expert Reasoning (KASSPER) dataset that contains heterogeneous data, array errors, and dense ground targets; 2) the Multi-Channel Airborne Radar Measurement (MCARM) dataset that was acquired from realistic airborne radar experiments; 3) the dataset for a bistatic radar geometry containing range-dependent clutter. Experimental results show that the parametric Rao and GLRT tests can provide good detection performance with limited or even no range training in more challenging environments. Therefore, these detectors offer useful solutions to detection problems in dense-target or heterogeneous environments.

## 7.2 Data Model and Problem Statement

Consider the problem of detecting a *known* multichannel signal with *unknown* amplitude in the presence of spatially and temporally correlated disturbance: (e.g., [1]):

$$\begin{aligned} H_0 : \quad & \mathbf{x}_0(n) = \mathbf{d}_0(n), \quad n = 0, \dots, N-1, \\ H_1 : \quad & \mathbf{x}_0(n) = \alpha \mathbf{s}(n) + \mathbf{d}_0(n), \quad n = 0, \dots, N-1, \end{aligned} \quad (7.1)$$

where all vectors are  $J \times 1$  vectors,  $J$  denotes the number of spatial channels, and  $N$  is the number of temporal observations. In the sequel,  $\mathbf{x}_0(n)$  is referred to as the *test signal*,  $\mathbf{s}(n)$  is the signal to be detected with amplitude  $\alpha$ , and  $\mathbf{d}_0(n)$  is the *disturbance signal* that may be correlated in space and time. In addition to the test signal  $\mathbf{x}_0(n)$ , there may be a set of *training* or *secondary* signals  $\mathbf{x}_k(n)$ ,  $k = 1, \dots, K$ , that are target-free:  $\mathbf{x}_k(n) = \mathbf{d}_k(n)$ .

In particular, we consider herein the signal detection problem in an airborne STAP radar system with  $J$  array channels and a coherent processing interval (CPI) of  $N$  pulses repetition intervals (PRIs). The disturbance  $\mathbf{d}_k(n)$ ,  $k = 0, \dots, K$  consists of ground clutter, jamming, and thermal noise. Let  $\mathbf{s} = [\mathbf{s}^T(0), \dots, \mathbf{s}^T(N-1)]^T$ . Similarly,  $\mathbf{d}_k$  and  $\mathbf{x}_k$  are formed from  $\mathbf{d}_k(n)$  and  $\mathbf{x}_k(n)$ , respectively. The target space-time steering vector  $\mathbf{s}$  is given by (assuming a uniform equi-distant linear array):

$$\mathbf{s}(\omega_s, \omega_d) = \mathbf{s}_s(\omega_s) \otimes \mathbf{s}_t(\omega_d), \quad (7.2)$$

where  $\mathbf{s}_s(\omega_s)$  and  $\mathbf{s}_t(\omega_d)$  denote the  $J \times 1$  spatial steering vector and the  $N \times 1$  temporal steering vector, respectively, and  $\omega_s$  and  $\omega_d$  denote the normalized target spatial and Doppler frequencies, respectively.

The binary composite hypothesis testing problem is to select between  $H_0 : \alpha = 0$  and  $H_1 : \alpha \neq 0$ . Classical STAP detectors (e.g., the RMB, Kelly's GLRT, and AMF detectors) were developed based on the assumption that [1, 9–13, 15, 16, 22, 23]:

- **AS1:** The disturbance signals  $\{\mathbf{d}_k\}_{k=0}^K$  are i.i.d. with distribution  $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ , where  $\mathbf{R}$  is the *unknown* space-time covariance matrix.

In addition to the above general assumption, a parametric model can be applied to characterize the disturbance [22, 23]:

- **AS2:** The disturbance signal  $\mathbf{d}_k(n)$ ,  $k = 0, \dots, K$ , can be modeled as a  $J$ -channel AR( $P$ ) process with *known* model order  $P$ :

$$\mathbf{d}_k(n) = -\sum_{i=1}^P \mathbf{A}^H(i) \mathbf{d}_k(n-i) + \boldsymbol{\varepsilon}_k(n), \quad (7.3)$$

where  $\{\mathbf{A}^H(i)\}_{i=1}^P$  denote the *unknown*  $J \times J$  AR coefficient matrices,  $\boldsymbol{\varepsilon}_k(n)$  denote the  $J \times 1$  spatial noise vectors that are temporally white but spatially colored Gaussian noise:  $\boldsymbol{\varepsilon}_k(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ , where  $\mathbf{Q}$  denotes the *unknown*  $J \times J$  spatial covariance matrix.

It is noted that assumption **AS1** is often violated in real airborne radar systems including *dense-target and heterogeneous* environments, which offer limited or even no range training data [19,21]. It is also shown that the performance of STAP detectors often degrades significantly in a heterogeneous environment where assumption **AS1** is violated because of the mismatch of the space-time covariance matrix relative to that of the target test cell. Moreover, the disturbances in real radar systems do not follow the multichannel AR model in **AS2**, whereas the parametric Rao and GLRT detectors assume that the disturbance follows the multichannel AR process. Therefore, the *problem of interest* is to evaluate the performance of the parametric Rao and GLRT detectors using more realistic data, i.e., the KASSPER, MCARM and Bistatic datasets.

### 7.3 Parametric Rao and GLRT Detectors

For easy reference and to facilitate our evaluation of the parametric detectors, we provide a brief summary of the parametric Rao and GLRT detectors in this section. These detectors are two different solutions to the problem stated in Section 7.2 [28,37,40,53]. The parametric Rao detector is computationally simpler, but the parametric GLRT offers improved performance. Both detectors first find the ML estimates of the unknown parameters, which are next used to compute the test statistics. The likelihood functions under the null and alternative hypotheses are parameterized by the signal amplitude  $\alpha$ , the AR coefficients  $\mathbf{A}^H = [\mathbf{A}^H(1), \dots, \mathbf{A}^H(P)] \in \mathbb{C}^{J \times JP}$ , and spatial covariance matrix  $\mathbf{Q}$ . Note that under the null hypothesis we have  $\alpha = 0$ . Given  $\mathbf{A}$ , the steering vector and test signal can be temporally whitened through the following inverse (i.e., moving average) filtering:

$$\tilde{\mathbf{s}}(n) = \mathbf{s}(n) + \sum_{i=1}^P \mathbf{A}^H(i) \mathbf{s}(n-i), \quad (7.4)$$

$$\tilde{\mathbf{x}}_0(n) = \mathbf{x}_0(n) + \sum_{i=1}^P \mathbf{A}^H(i) \mathbf{x}_0(n-i). \quad (7.5)$$

This is an important observation exploited by the parametric Rao and parametric GLRT detectors that are summarized next.

The parametric GLRT is given by [28,53]

$$T_{\text{GLRT}} = 2L \ln \frac{|\hat{\mathbf{Q}}_{\text{ML},0}|}{|\hat{\mathbf{Q}}_{\text{ML},1}|} \underset{H_0}{\overset{H_1}{\geq}} \gamma_{\text{GLRT}}, \quad (7.6)$$

where  $L = (K + 1)(N - P)$  and  $\gamma_{\text{GLRT}}$  denotes the corresponding test threshold. The ML estimates of the spatial covariance matrix under the null and alternative hypotheses,  $\hat{\mathbf{Q}}_{\text{ML},0}$  and  $\hat{\mathbf{Q}}_{\text{ML},1}$  are given by

$$\hat{\mathbf{Q}}_{\text{ML},0} = \hat{\mathbf{Q}}(\alpha)|_{\alpha=0}, \quad (7.7)$$

$$\hat{\mathbf{Q}}_{\text{ML},1} = \hat{\mathbf{Q}}(\alpha)|_{\alpha=\hat{\alpha}_{\text{ML}}}. \quad (7.8)$$

The  $\alpha$ -dependent  $\hat{\mathbf{Q}}(\alpha)$  is given by

$$\hat{\mathbf{Q}}(\alpha) = \frac{1}{L} \left( \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha) \right), \quad (7.9)$$

where the  $\alpha$ -dependent correlation matrices are

$$\begin{aligned} \hat{\mathbf{R}}_{xx}(\alpha) &= \sum_{k=1}^K \sum_{n=P}^{N-1} \mathbf{x}_k(n) \mathbf{x}_k^H(n) \\ &\quad + \sum_{n=P}^{N-1} [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)] [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)]^H, \end{aligned} \quad (7.10)$$

$$\begin{aligned} \hat{\mathbf{R}}_{yy}(\alpha) &= \sum_{k=1}^K \sum_{n=P}^{N-1} \mathbf{y}_k(n) \mathbf{y}_k^H(n) \\ &\quad + \sum_{n=P}^{N-1} [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)] [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)]^H, \end{aligned} \quad (7.11)$$

$$\begin{aligned} \hat{\mathbf{R}}_{yx}(\alpha) &= \sum_{k=1}^K \sum_{n=P}^{N-1} \mathbf{y}_k(n) \mathbf{x}_k^H(n) \\ &\quad + \sum_{n=P}^{N-1} [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)] [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)]^H, \end{aligned} \quad (7.12)$$

with  $\mathbf{t}(n)$  and  $\mathbf{y}_k(n)$  denoting the regression subvectors formed from the steering vector  $\mathbf{s}(n)$  and test signal  $\mathbf{x}_k(n)$ , respectively:  $\mathbf{t}(n) = [\mathbf{s}^T(n-1), \dots, \mathbf{s}^T(n-P)]^T \in \mathbb{C}^{JP \times 1}$  and  $\mathbf{y}_k(n) = [\mathbf{x}_k^T(n-1), \dots, \mathbf{x}_k^T(n-P)]^T \in \mathbb{C}^{JP \times 1}, k = 0, \dots, K$ . The ML estimate of  $\alpha$  under the alternative hypothesis, which is used in (7.8), is given by

$$\hat{\alpha}_{\text{ML}} = \arg \min_{\alpha} \left| \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha) \right|. \quad (7.13)$$

The parametric Rao test is given by [37, 40]

$$T_{\text{Rao}} = \frac{2 \left| \sum_{n=P}^{N-1} \hat{\mathbf{s}}^H(n) \hat{\mathbf{Q}}_{\text{ML},0}^{-1} \hat{\mathbf{x}}_0^H(n) \right|^2}{\sum_{n=P}^{N-1} \hat{\mathbf{s}}^H(n) \hat{\mathbf{Q}}_{\text{ML},0}^{-1} \hat{\mathbf{s}}^H(n)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{Rao}}, \quad (7.14)$$

where  $\gamma_{\text{Rao}}$  denotes the test threshold. The temporally whitened steering vector  $\hat{\mathbf{s}}^H(n)$  and test signal  $\hat{\mathbf{x}}_0^H(n)$  are obtained by replacing  $\mathbf{A}^H$  with its ML estimate under  $H_0$

$$\hat{\mathbf{A}}_{\text{ML},0}^H = -\hat{\mathbf{R}}_{yx}^H(\alpha)\hat{\mathbf{R}}_{yy}^{-1}(\alpha)|_{\alpha=0}, \quad (7.15)$$

in (7.4) and (7.5), respectively.

The flowcharts for parametric GLRT and Rao detectors are shown in Figs. (7.1) and (7.2). The Rao test is shown to be asymptotically equivalent to the GLRT but may be inferior to the latter when the data size is small. In addition, the Rao test is obtained based upon a low-order Taylor expansion of the GLRT, an approximation which is only valid for weak signals [7]. As such, the performance of the parametric Rao detector degrades when the weak signal assumption is violated. The parametric GLRT was developed as an improved detector to deal with the above issues. However, the cost function of the ML amplitude estimator in (7.13) is highly non-linear. Newton-like iterative non-linear searches are generally used to find the ML amplitude estimate. Another sub-optimum but computationally more efficient estimator, referred to as the asymptotic ML (AML) estimator, was developed in [28, 53]. The AML estimator, which was found to yield similar performance to the ML estimator, can be implemented as follows:

- **Step 1** First, compute a least-squares (LS) amplitude estimate  $\hat{\alpha}_{\text{LS}} = \frac{\mathbf{s}^H \mathbf{x}_0}{\mathbf{s}^H \mathbf{s}}$ . Then, determine an estimate  $\hat{\mathbf{A}}_{\text{LS}}^H$  of  $\mathbf{A}^H$  as follows:

$$\hat{\mathbf{A}}_{\text{LS}}^H = \hat{\mathbf{R}}_{yx}^H(\hat{\alpha}_{\text{LS}})\hat{\mathbf{R}}_{yy}^{-1}(\hat{\alpha}_{\text{LS}}), \quad (7.16)$$

which can be shown to be statistically consistent [28, 53].

- **Step 2** Compute the temporally whitened signals  $\hat{\mathbf{x}}_k(n)$  and  $\hat{\mathbf{s}}(n)$  by replacing  $\mathbf{A}^H$  with the LS AR coefficient estimate  $\hat{\mathbf{A}}_{\text{LS}}^H$  in (7.4) and (7.5), respectively. Then, obtain the AML amplitude estimate  $\hat{\alpha}_{\text{AML}}$  by using

$$\hat{\alpha}_{\text{AML}} = \frac{\text{tr}(\hat{\mathbf{S}}^H \boldsymbol{\Psi}^{-1} \hat{\mathbf{X}}_0)}{\text{tr}(\hat{\mathbf{S}}^H \boldsymbol{\Psi}^{-1} \hat{\mathbf{S}})}, \quad (7.17)$$

where  $\hat{\mathbf{S}} = [\hat{\mathbf{s}}(P), \dots, \hat{\mathbf{s}}(N-1)] \in \mathbb{C}^{J \times (N-P)}$ ,  $\hat{\mathbf{X}}_k = [\hat{\mathbf{x}}_k(P), \dots, \hat{\mathbf{x}}_k(N-1)] \in \mathbb{C}^{J \times (N-P)}$ , and

$$\boldsymbol{\Psi} = \hat{\mathbf{X}}_0 \mathbf{P}^\perp \hat{\mathbf{X}}_0^H + \sum_{k=1}^K \hat{\mathbf{X}}_k \hat{\mathbf{X}}_k^H, \quad (7.18)$$

with  $\mathbf{P}^\perp$  denoting the projection matrix projecting to the orthogonal complement of the range of  $\hat{\mathbf{S}}^H$ :  $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P} = \mathbf{I} - \hat{\mathbf{S}}^H (\hat{\mathbf{S}}^H)^\dagger \in \mathbb{C}^{(N-P) \times (N-P)}$ .

- **Step 3** Find the AML estimate of the spatial covariance matrix by substituting  $\hat{\alpha}_{\text{AML}}$  for  $\alpha$  in (7.9).

Recall that the parametric Rao and parametric GLRT detectors utilize both the test and training signals for the parameter estimation. As a result, they are functional even without range training data [28, 53]. The capability to handle the training-free detection is a unique and desirable attribute of the parametric detectors which is not shared by other existing detectors including the PAMF detector.

## 7.4 Performance Evaluation

This section is to test the parametric detectors using the KASSPER [52], MCARM and bistatic datasets. The performance evaluation of the parametric detectors with both datasets allows us to assess the influence from the mismatch between **AS2** and real disturbances.

### 7.4.1 Simulated KASSPER Dataset

The KASSPER 2002 dataset contains many real-world effects including heterogeneous terrain, subspace leakage, array errors, and many ground targets. The simulated airborne radar was flying at 3000 m altitude at 100 m/s traveling due east with a  $3^\circ$  crab angle. The radar was operating at 1240 MHz with a peak power of 15 kW. The 11 (virtual) antenna array elements were spaced slightly less than a half-wavelength apart at 0.1092 m (0.9028 half-wavelength spacing), and the transmit array is uniformly weighted in the horizontal dimension and phased to steer the mainbeam to  $195^\circ$ . The pulse repetition frequency (PRF) was 1984 Hz and the CPI contains 32 pulses.

#### Dense Targets Environment

The KASSPER dataset simulates dense targets environment. Of particular interest are the targets in the mainbeam of the radar within the range swath of interest. In total, there are 268 targets from 35 to 50 km.

We apply the parametric detectors to the KASSPER data of interest and compute the test statistics with respect to Doppler frequency and range cells. To compare with other techniques, we count the number of detection while setting the threshold with constraint on the number of false alarm. Due to range/Doppler sidelobes resulting from pulse compression and Doppler filtering, it is common for a target to spread into nearby range-Doppler cells. For this reason, it is standard procedure for a radar to cluster target detections such that a detection in a given range-Doppler cell is associated with a target that lies in a contiguous range cell or Doppler band. For the results reported here, we cluster  $\pm 1$  cells in range and  $\pm 1$  bands in Doppler for all processing schemes. Once we declare a detection in its region, corresponding test value is removed from original test statistics to avoid over-counting detection.

For comparison purpose, the Joint Domain Localized (JDL) technique with  $3 \times 3$  local processing region (LPR) is applied to the KASSPER dataset. There are 64 Doppler frequency bands and two guard cells are used at each side of the test range cell. The number of false alarm  $N_{fa}$

Table 7.1: Number of Detected Targets when the number of false alarm  $N_{fa} = 10$  for  $J = 11$  without training limiting and with IPS training limiting (shown in the parenthesis).

	Parametric GLRT	Parametric Rao	JDL
$K = 11, P = 1$	37 (72)	33 (59)	19 (35)
$K = 11, P = 2$	44 (67)	31 (46)	19 (35)
$K = 22, P = 1$	42 (73)	43 (67)	20 (38)
$K = 22, P = 2$	49 (101)	41 (76)	20 (38)

is constrained to 10. For the parametric detectors and JDL, two cases are considered: without training limiting and with IPS training limiting.

**No training limiting:** Table 7.1 lists the number of detection when the number of false alarm  $N_{fa} = 10$  for  $J = 11$ , where true targets are represented by cross signs and detections by bars. From this table, having more training data does not improve detection performance without training selection. The clairvoyant detector with known covariance matrices can detect 176 out of the 268 targets and the SMI with  $K = 999$  training data can detect 32 targets. The JDL can detect up to 20 targets, while the parametric detectors can detect at most 49 targets.

**IPS training limiting:** To improve the detection performance, we limit the training data by applying the Innovation Power Sorting (IPS). In this case, the performance for each technique is listed in Table 7.1, as shown in the parenthesis. The results show that, with IPS training selection, the performance of the parametric detectors improves considerably. When  $J = 11$ ,  $K = 22$  and  $P = 2$ , the parameter GLRT detectors can detect over 100 targets, while the JDL can detect 38 targets.

## Heterogeneous Environment

Since the KASSPER 2002 dataset contains the true interference covariance matrix for every range bin, we can realize the disturbances by following Gaussian processes with the true covariance matrices, and obtain the detection performance of the parametric detectors with respect to SINR ( $P_d$  versus SINR for a fixed  $P_f$ ). Meanwhile, we have shown that having large ratio of the number of pulses to the number of channels can improve the performance in case of limited training [28]. Since the number of channel in KASSPER dataset is fixed as  $N = 32$ , we alternatively generate the CPI data for  $J = 4$ . Since the disturbance signals generated as above do not follow an AR model in general and are more realistic than our earlier simulated data using an AR model, it would be of interest to determine the detection performance of the parametric detectors using this simulated data set and compare with our earlier simulation results based on an AR model in [28, 40]. Besides the violation of **AS2**, we investigate the performance of the parametric detectors in a simulated *heterogenous environment*, which does not assume i.i.d. in **AS1**. To this end, we generate training signals using different covariance matrices from that of the range bin under test.

In this simulation, we select the range bin 200 as the testing cell. The SINR is defined as

$$\text{SINR} = |\alpha|^2 \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s} \quad (7.19)$$



where the space-time covariance matrix  $\mathbf{R}$  is chosen by loading the *Rcc\_r200.mat* in the KASSPER 2002 dataset.

Figure 7.3 presents the estimated angle-Doppler power spectral density (PSD) of the disturbance in range bin 200 when  $J = 11$  and  $N = 32$ . It shows that there exists a high clutter centered in the normalized Doppler frequency of 0.1 and azimuth angle of  $195^\circ$ . Figures 7.4(a) and 7.4(b) depict the probability of detection versus SINR with  $K = 0$  (no training) or  $K = 1$  (limited training), and  $P = 1$ . For the no training case, the performance of both parametric detectors degrades, while the performance is improved with one range training data. It is worthy noting that, for the limited training case  $K = 1$ , the training data was generated by using different covariance matrix from the covariance matrix for the test data, which may be true in the heterogeneous environments.

In [28,40], we have shown that the performance degradation caused by the absence of training can be mitigated by using a larger  $N$ , i.e., increasing temporal observations of the test signal. Since  $N = 32$  is fixed in the KASSPER 2002 dataset, we alternatively down-sample the CPI datacube to increase the ratio of  $N$  to  $J$ , i.e., extracting data corresponding to  $J = 4$  channels from the original dataset.

Figure 7.4(c) shows the case with  $J = 4$  and  $N = 32$  with no training data, and Figure 7.4(d) shows the cases with limited training. It is seen that having larger ratio of  $N$  to  $J$  improves the detection performance. Specifically, with  $K = 1$ , the performance of the parametric detectors is only about 3 dB inferior to the ideal MF detector with known covariance matrix.

## 7.4.2 Measured MCARM Dataset

In this part, the MCARM dataset, the real world multi-channel airborne data containing clutter in various terrains including mountains, rural, urban, and land/sea interface, is considered. The MCARM data is collected from the BAC 1-11 airborne platform in the L-Band frequency. The MCARM array has 16 columns, each consisting of two four-element subarrays. Each subarray has its own output or is combined into a single output per column with up to 24 outputs for the array.

Since the true joint space-time covariance matrix of MCARM dataset is unavailable, another power measure is adopted since it can be estimated from the data. It is the *input SINR* (per-pulse, per-channel) defined as

$$\text{SINR}_{\text{IN}} = \frac{|\alpha|^2}{\sigma_d^2}, \quad (7.20)$$

where  $\alpha$  is the target amplitude and  $\sigma_d^2$  denotes the variance (power) of each element of the disturbance vector at each time instant. We carried out the performance analysis of the parametric Rao and GLRT detectors using two acquisitions from the MCARM database. More specifically, rd050575 and rel050152 have been used to assess the performance of the parametric detectors.

### Case I - Inserted Artificial Target

In this case, we choose the dataset according to the following parameters:

- One elevation angle ( $0^\circ$ )
- One azimuth angle ( $0^\circ$ )
- Four channels ( $J = 4$ )
- Various  $K$  (training),  $N$  (pulses), and normalized Doppler frequencies

An artificial target with an  $\text{SINR}_{\text{IN}}$  of  $-30\text{dB}$  is injected in the range bin 295. The disturbance power  $\sigma_d^2$  is estimated as a five-bin average centered on the range bin in which the target is placed. Model order values  $P = 1, 2, 3, 4$  were evaluated for each parametric tests and the model order with the best performance was selected. The selection criterion is the difference between the target peak value and the highest non-target peak value. Diagonal loading of 40 dB for the AMF detector is applied.

Figure 7.5 depicts the estimated angle-Doppler power spectral density of the disturbance in RB 295 when  $J = 4$  and  $N = 128$ . It shows that there exists a high clutter centered in the normalized Doppler frequency of  $-0.05$  and azimuth angle of  $0^\circ$ .

Figure 7.6(a) shows the case where the target is located outside the clutter (the normalized Doppler frequency is 0.2). However, if the target is located inside the clutter (the normalized Doppler frequency is 0.1) the AMF detector does not work at all, but the parametric Rao and GLRT detectors work well (See Figures 7.6(c) and 7.6(d)).

Fig. 7.6(c) shows the test statistics for the parametric Rao and GLRT detectors without training support, the AMF with  $K = 8$  and the JDL with  $K = 8$ , when the target is located inside the clutter. Specifically, the AMF and JDL cannot work without training data. Clearly, the parametric Rao and GLRT detectors can detect the inserted target with over 20 dB stronger peak. Fig. 7.6(d) shows the test statistics for the case of  $K = 4$ .

## Case II - Five Targets

In this case, no artificial target was injected. File re050152 employs a moving target simulator in the radar field of view to simulate a moving point source. The dataset consists of  $N = 128$  pulses,  $J = 22$  channels and 630 range cells. There are five targets in total. The angle-Doppler power spectral density for range bin 450 is shown in Fig. 7.7, when  $J = 22$  and  $N = 128$ . It shows that there exist five targets equi-spaced in normalized Doppler frequency, and centered azimuth angle of  $-15^\circ$ .

It shows that there exist five targets equi-spaced in normalized Doppler frequency, and centered azimuth angle of  $-15^\circ$ . Figs. 7.8(a) and 7.8(b) depict the test statistics for the parametric Rao and AMF detectors when  $J = 22$ ,  $N = 128$ ,  $P = 1$  and  $K = 11$  or  $K = 22$ . It is seen that the parametric Rao detector detects all five targets, while the JDL can detect up to four targets and the AMF detects two targets.

Table 7.2: The Operating Parameters for the Bistatic Dataset

Parameter	Value
carrier frequency	1240 MHz
receiver bandwidth	1 MHz
number of pulses	64
number of channels	16
number of range bins	601
pulse repetition frequency	400 Hz
peak transmit power	19.2 kW
transmitting platform altitude	3.1 km
transmitting platform speed	0 m/s (stationary) 100 m/s (moving)
receiving platform altitude	4 km
transmitting platform speed	100 m/s
tilt angle (Tx/Rx)	5°

### 7.4.3 Bistatic Dataset

The Bistatic dataset is simulated bistatic airborne radar data contains non i.i.d. and range-dependent clutter due to bistatic geometry. The bistatic data used in the performance evaluation of the proposed algorithms was generated using Stiefvater Consultants Signal Modeling and Simulation (SMS) tool [57] and follows closely results discussed in [58]. Two cases have been considered. In both cases, the receiver is assumed to be moving at a velocity of 100 m/sec while the transmitter's velocity is assumed to be 0 m/sec (case I) and 100 m/sec (case II), with an offset angle of  $45^\circ$ . The transmitter center frequency is  $f_c = 1.24\text{GHz}$ , the receiver height is  $h_R = 3.1\text{km}$ , the transmitter height is  $h_T = 4\text{km}$ , and the baseline separation is  $L = 100\text{km}$ . The pulse repetition frequency (PRF) was assumed to be 400 Hz. In both cases, the test cell is located at 35 km from the receiver and the azimuth of the test cell with respect to the receiver is  $135^\circ$ , producing a bistatic angle of  $33.9^\circ$ . The test cell is located at 1m above the ground. In each case, a linear array composed of  $J = 32$  sensors and  $N = 64$  pulses have been used. Data for 601 range cells was generated using the same radar system parameters as those used in the MCARM experiment [59]. The bistatic geometry is shown in Fig. 7.9.

In both cases there is only one target in the test cell 301. The operating parameters for the simulated Bistatic dataset are shown in Table 7.2.

#### Case I - Non i.i.d Case

In the case of a stationary transmitter, note from Fig. 7.10(a) that unlike the monostatic side-looking airborne radar, where all traces are overlapping and the clutter spectral centers co-located, the clutter spectral centers are spread both in angle and Doppler. This generates spectral dispersion, thus making the secondary data vary with range. This in turn makes the training data *non independent and identically distributed (i.i.d)*, hence violating a fundamental assumption made

Table 7.3: Case I: Difference between the target peak and the highest non-target peak

	P-GLRT/ML	P-Rao	P-GLRT/AML2	JDL
$K = 0$	<b>2.96 dB</b>	0.66 dB	2.51 dB	-
$K = 2$	9.50 dB	8.17 dB	<b>10.27 dB</b>	-
$K = 4$	11.58 dB	9.76 dB	<b>13.63 dB</b>	-
$K = 8$	14.31 dB	12.57 dB	<b>16.27 dB</b>	15.24 dB
$K = 16$	17.25 dB	15.64 dB	<b>19.08 dB</b>	17.29 dB

in several STAP algorithms.

For case I, Figure 7.11 shows the test statistic of the parametric Rao detector, the parametric GLRT detector, and the joint domain localized (JDL) with  $3 \times 3$  local processing region (LPR) with respect to normalized Doppler frequency and range bins. The detection of the target is marked by the rectangle. The number of training for the parametric detectors is  $K = 8$ . For the JDL, when the number of training is limited to  $K = 8$  which does not satisfy the full-rank estimate of the covariance matrix in the angle-Doppler domain, the matrix pseudo-inversion is used. To facilitate performance comparison, the differences between the target peak and the highest non-target peak for various detectors are shown in the Tables 7.3. Clearly, the parametric Rao and GLRT detectors can detect a target with limited or even no training data support. Specifically, the parametric GLRT detector can detect a target without training data support with a target detection 2.96 dB higher than other test statistics. When the number of training is increased, the performance of all detectors improves. For  $K = 18$ , all three detectors exhibit similar detection performance.

## Case II - Range-dependent Case

In the case of a moving transmitter, it is clear from Fig. 7.10(b) that the angle-Doppler traces are non-overlapping and the spectral centers are highly distributed in angle and Doppler, due to the relative motion of the receive and transmit platforms. Significant spectrum dispersion over range then arises, which leads to significant performance degradation [58].

For case II, Figure 7.12 shows the test statistic of the parametric detectors and the JDL with respect to normalized Doppler frequency and range bins. The number of training is specified in the figures. In extremely case, i.e.,  $K = 0$ , the parametric Rao and GLRT detectors have 1.84 dB and 4.47 dB stronger output, by comparing the test statistics at range bin 301 with ones at other range bins. When  $K = 8$ , the JDL cannot find any target, while the parametric Rao and GLRT detectors can detect the true target at range bin 301 with 12.94 dB and 14 dB stronger amplitude, respectively. By increasing the number of training to  $K = 16$ , the parametric Rao and GLRT detectors exhibit a 13.92 dB and 14.59 dB higher peak at range bin 301, while the JDL finds two distinct peaks which are located at range bin 100 and 301, respectively. The differences between the target peak and the highest non-target peak for various detectors in this case are shown in the Table 7.4, where the symbol – represents the detector fails to find the target, and negative value means the wrong detection has higher peak than the correct detection.

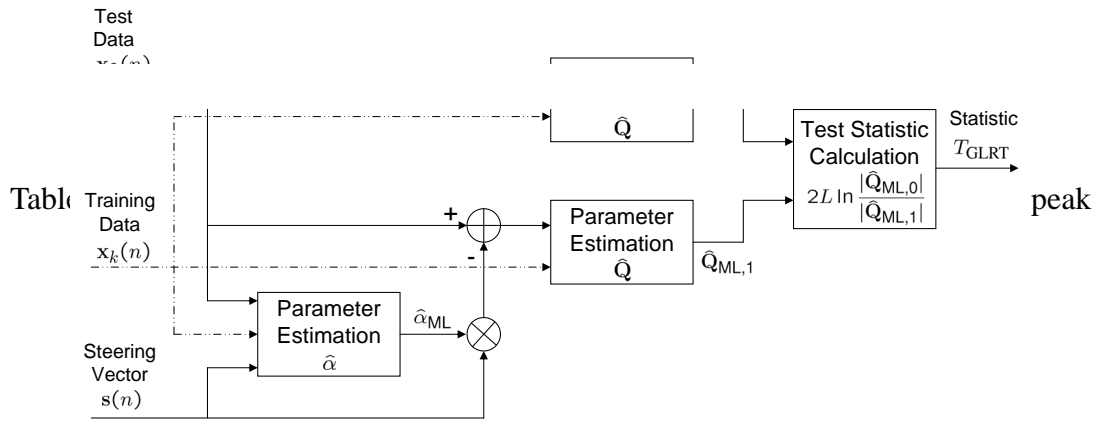


Figure 7.1: The flowchart for the parametric GLRT detector

## 7.5 Conclusions

We have examined the performance of the parametric Rao and GLRT detectors using the KASSPER 2002, MCARM and Bistatic datasets include many real-world effects. Our results show that these parametric detectors work quite well with limited or no range training data support in more realistic environments. Therefore, they are good candidates for solving detection problems in the presence of range dependent clutter and/or in heterogeneous environments.

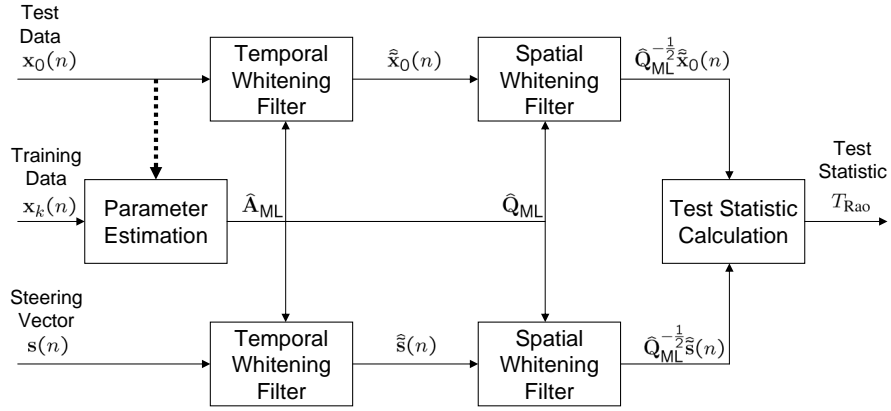


Figure 7.2: The flowchart for the parametric Rao detector

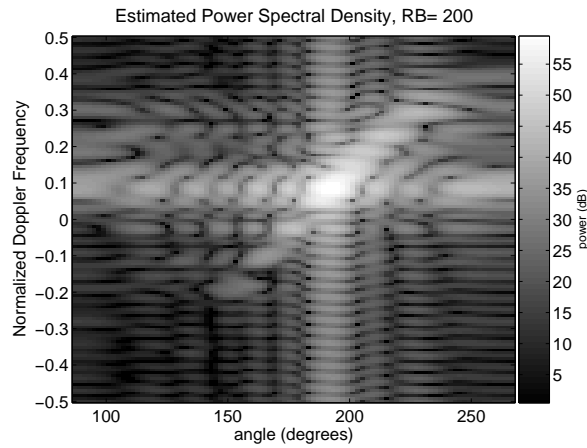


Figure 7.3: The angle-Doppler power spectral density of the range bin 200 when  $J = 11$  and  $N = 32$

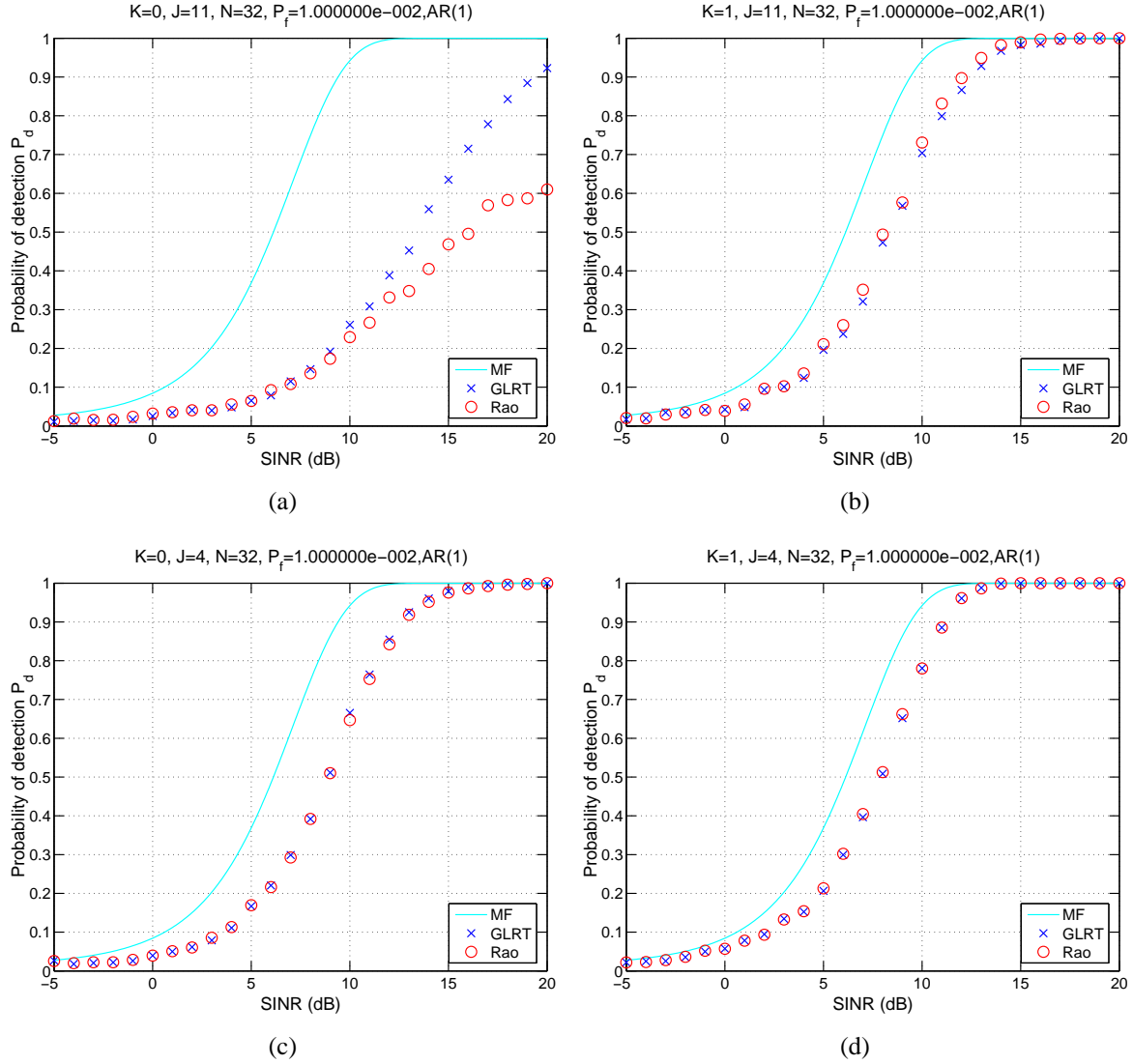


Figure 7.4: The probability of detection versus input SINR for the KASSPER 2002 dataset.

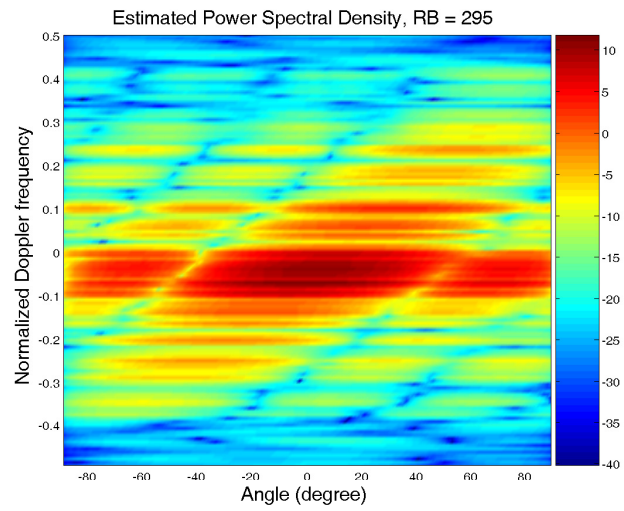


Figure 7.5: Estimated angle-Doppler power spectral density for range bin 295 when  $J = 4$  and  $N = 128$ .



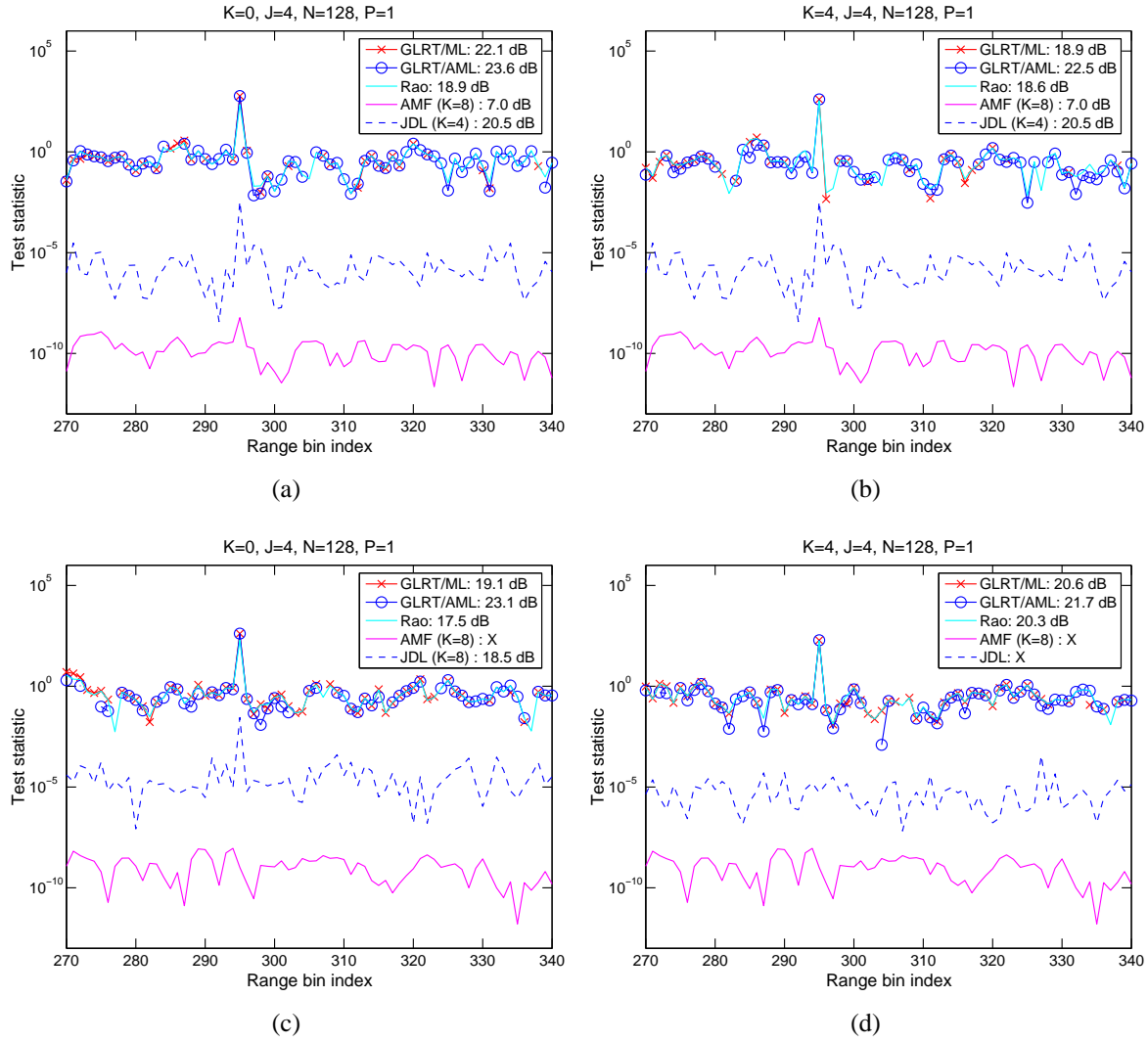


Figure 7.6: The probability of detection versus input SINR for the KASSPER 2002 dataset.

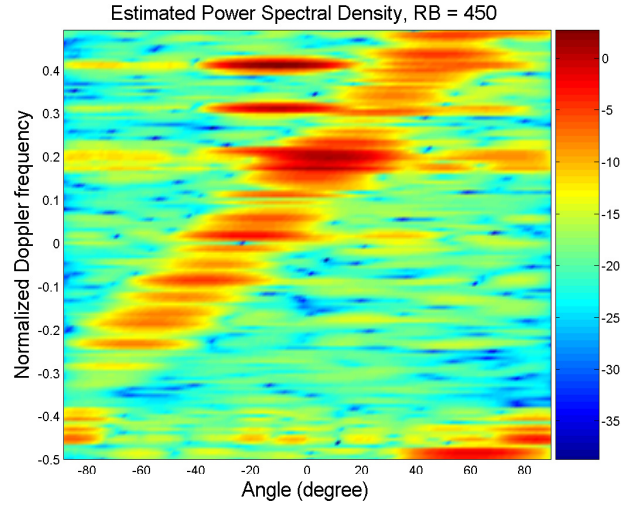


Figure 7.7: Estimated angle-Doppler power spectral density for range bin 450 when  $J = 22$  and  $N = 128$ .

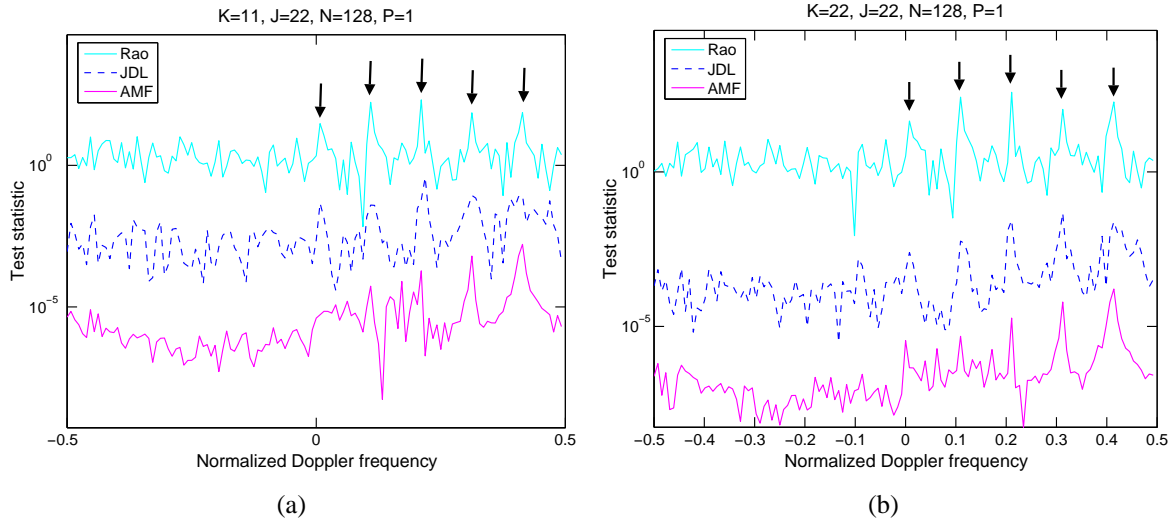


Figure 7.8: Test statistics of the parametric Rao, JDL and AMF detectors when  $J = 22$ ,  $N = 128$ ,  $P = 1$  and  $K = 11$  (a)  $K = 22$  (b).

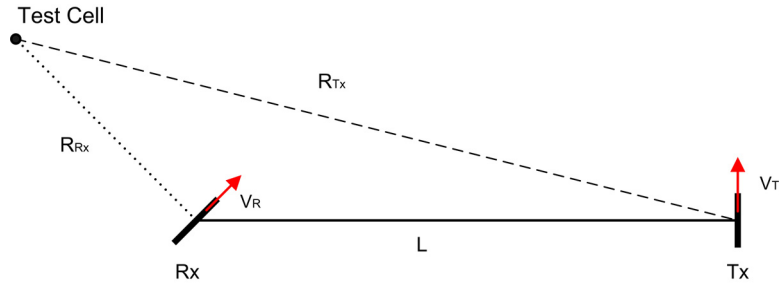


Figure 7.9: Bistatic airborne radar geometry.

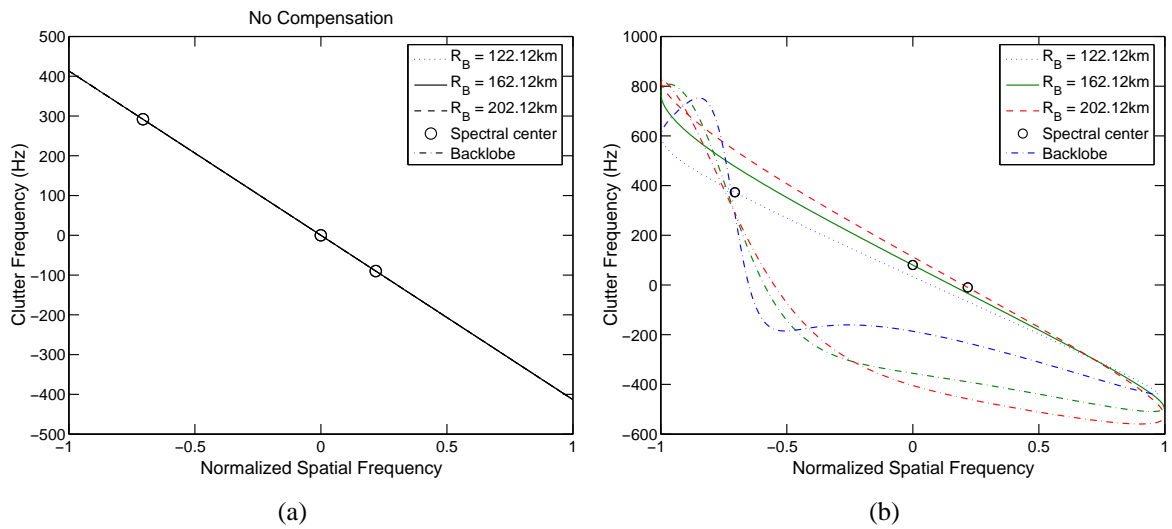


Figure 7.10: Bistatic Angle-Doppler traces, Case I (a) and Case II (b).

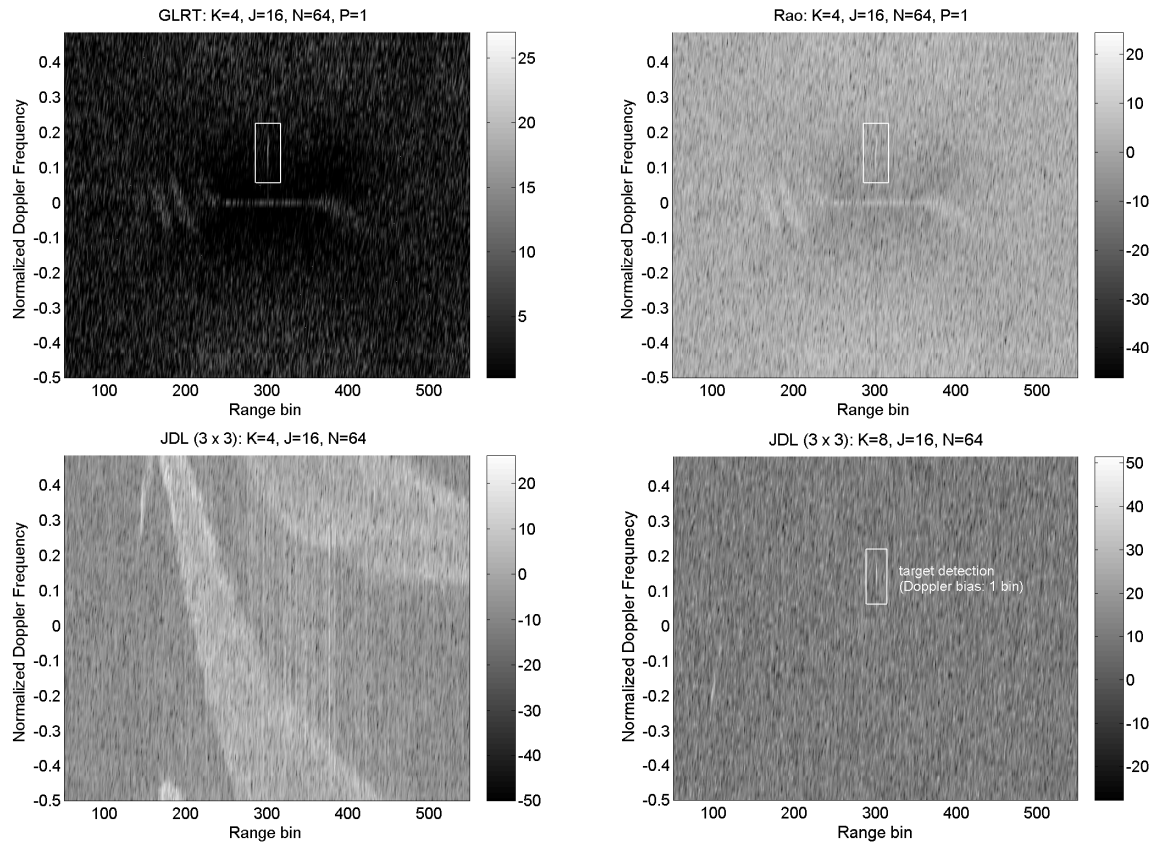


Figure 7.11: Case I: Test statistics of parametric GLRT, parametric Rao and JDL with respect to normalized Doppler frequency and range bins when  $J = 16$  and  $N = 64$ . The detection is marked by the rectangle.

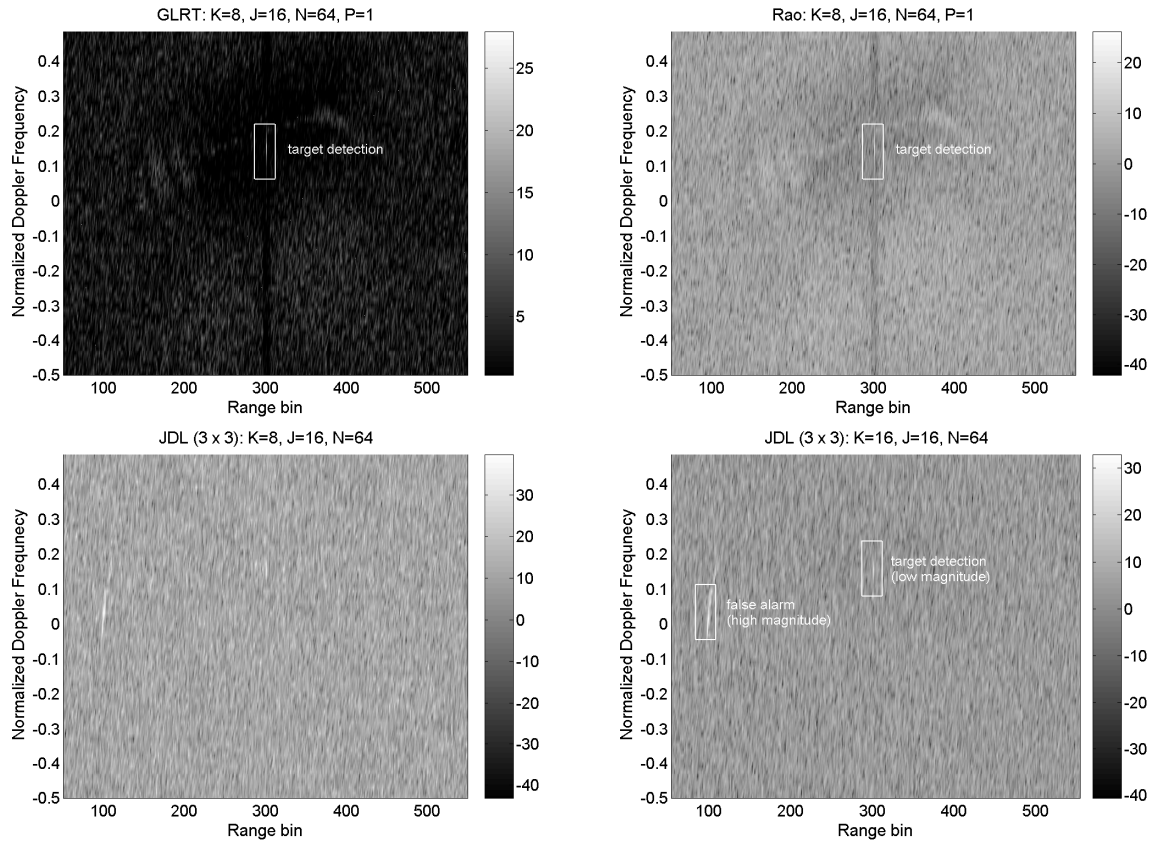


Figure 7.12: Case II: Test statistics of parametric GLRT, parametric Rao and JDL with respect to normalized Doppler frequency and range bins when  $J = 16$ ,  $N = 64$ . The detection is marked by the rectangle. The JDL with  $K = 8$  fails to detect any target, while the JDL with  $K = 16$  has a strong detection around range bin 100 and a weak detection at range bin 301.

## Chapter 8

# Parametric Adaptive Signal Detection for Hyperspectral Imaging

### 8.1 Introduction

Hyperspectral sensors are a new class of imaging spectroscopy sensors that divide the waveband of interest into hundreds of contiguous narrow bands. Their fine spectral resolution enables remote identification of ground objects based on their spectral signatures. Hyperspectral imaging (HSI) has a wide range of applications including terrain classification, environmental and agricultural monitoring, geological exploration, ordinance remediation, tactical surveillance, and others [60].

A challenging problem in HSI applications is the so-called *subpixel target detection*, which involves detecting objects occupying only a portion of a full pixel in an HSI image [4]. In such a case, the signal produced by the HSI sensors consists of both the *object* and *background*, the latter behaving effectively as interference that has to be suppressed for effective detection. The problem is reminiscent of that of detecting a known signal with unknown amplitude in colored noise with unknown correlation<sup>1</sup> (e.g., [7]). A multitude of solutions have been developed, including the Kelly's generalized likelihood ratio (GLR) test [13], adaptive matched filter (AMF) [12], adaptive coherence estimator (ACE) test [14, 15], among others. While these detectors can be used to solve the HSI subpixel target detection problem, there is a major difficulty with them in training-limited scenarios. In particular, the above detectors are *covariance-matrix based techniques* in that they all rely on an estimate of the background covariance matrix, which is obtained from *target-free* training pixels. The size of the background covariance matrix is identical to the number of spectral bands that is typically in the order of hundreds. A good estimate of the covariance matrix would require several hundred or more target-free training pixels, which may not be available in heterogeneous or dense-target environments. Another problem with the above covariance-matrix based detectors is complexity, since the large-size covariance matrix has to be estimated and

---

<sup>1</sup>We take a stochastic approach herein by modeling the background as a correlated random vector with an unknown covariance matrix. There are other detectors based on modeling the background as a deterministic quantity. See [4, 61, 62] and references therein for details.

inverted frequently.

There is a significant interest in developing *training-efficient* detection techniques for training-limited applications, such as the above HSI target detection approaches applied in heterogeneous environments. Another example is target detection based on space-time adaptive processing (STAP) for airborne radars [1], where range-dependent clutter characteristics, along with other issues, prevent inclusion of a large number of range cells far away from the test cell in the training set. One effective way to reduce training requirement in STAP detection is to utilize a suitable parametric model for the radar clutter and exploit the model for target detection. In particular, multichannel autoregressive (AR) models have been found to be very effective in representing the temporal correlation among pulse returns [22]- [25]. A parametric detector based on such a multichannel AR clutter model is developed in [22]- [23], which is referred to as the parametric adaptive matched filter (PAMF). The PAMF detector has been shown to significantly outperform the covariance-matrix based detectors for small training size.

For HSI applications, however, the data is non-stationary in the spectral domain (see Section 8.4.1 for details of such non-stationarity),<sup>2</sup> whereas AR models are by definition stationary. To account for such non-stationarity, we introduce in this chapter a sliding-window based non-stationary AR (NS-AR) model to capture the spectral correlation of HSI data. We propose a class of parametric adaptive signal detectors for HSI subpixel target detection, and develop a maximum likelihood (ML) estimation algorithm to estimate the parameters associated with the NS-AR model. In addition, we develop model order selection, training screening, and time-series based whitening and detection techniques, which are intrinsic parts of the proposed parametric adaptive detectors. We show via experimental results with real HSI data that our proposed parametric detectors are more efficient in training usage and outperform the conventional covariance-matrix based detectors when the training size is limited.

The rest of the chapter is organized as follows. Section 8.2 contains the data model and problem statement. The covariance-matrix based detectors are briefly reviewed and discussed in Section 8.3. The proposed techniques, including an NS-AR model, a class of parametric adaptive detectors, an ML parameter estimation algorithm, a model order selection method, and a training screening approach, are detailed in Section 8.4. Experimental results illustrating the performance of the proposed detectors under homogeneous, heterogeneous, and dense-target environments are presented in Section 8.5. Finally, Section 8.6 contains our concluding remarks.

## 8.2 Data Model and Problem Statement

Obtained through both spatial and spectral sampling, HSI data is usually described as a datacube, whose face is a function of the spatial coordinates and depth is a function of spectral bands or wavelengths. Each pixel can be represented as an  $L \times 1$  *real-valued* vector:  $\mathbf{x} = [x(0), x(1), \dots, x(L-1)]^T$ , where  $L$  denotes the total number of spectral bands,  $x(l)$  denotes the spectral response at the  $l$ th spectral band, and  $(\cdot)^T$  denotes transpose. Since HSI data

---

<sup>2</sup>Such spectral non-stationarity shall not be confused with the spatial stationarity which is often assumed for HSI data [4].

has non-zero mean [4, 65], a preprocessing stage is usually invoked to remove the sample mean estimated using the neighbor pixels.

In vector notation, the *subpixel signal detection* problem is described by the following composite hypothesis test [4]:

$$\begin{aligned} H_0 : \quad \mathbf{x} &= \mathbf{b}, & \text{target absent} \\ H_1 : \quad \mathbf{x} &= a\mathbf{s} + \mathbf{b}, & \text{target present} \end{aligned} \quad (8.1)$$

where  $\mathbf{x} \in \mathbb{R}^{L \times 1}$  is the demeaned test pixel,  $\mathbf{s} \in \mathbb{R}^{L \times 1}$  is the *signature vector* of the target object with amplitude  $a$ , and  $\mathbf{b} \in \mathbb{R}^{L \times 1}$  denotes the background plus system noise. We adopt the standard assumption that the signature vector  $\mathbf{s}$  is deterministic and known to the detector;<sup>3</sup> the amplitude  $a$ , however, is assumed unknown. For the background, we follow a *statistical approach* that models the background interference  $\mathbf{b}$  as a multivariate Gaussian random vector with zero mean and an unknown covariance matrix  $\mathbf{R}_b = E\{\mathbf{b}\mathbf{b}^T\}$ . The Gaussian assumption has been widely used for multispectral (e.g., [65]) and HSI data [4]. It leads to mathematical tractability and reasonably good performance. Nevertheless, it should be noted that a Gaussian model is not fully appropriate to characterize the statistical behavior of HSI data in many realistic cases, and alternative modeling approaches have been considered in [66]- [68].

Equation (8.1) implies that the background interference covariance matrix is the same under both hypotheses. Since for a subpixel target the area covered by background is different under the two hypotheses, it is more appropriate to consider the following modified hypothesis [4, 69]:

$$\begin{aligned} H_0 : \quad \mathbf{x} &= \mathbf{b}, & \text{target absent} \\ H_1 : \quad \mathbf{x} &= a\mathbf{s} + \sigma\mathbf{b}, & \text{target present,} \end{aligned} \quad (8.2)$$

where  $\sigma$  is unknown and, along with the signature amplitude  $a$ , determined by the target *fill factor*, i.e., the percentage of the pixel area occupied by the target [4].

Similar to [4], we assume that in addition to the test pixel  $\mathbf{x}$ , we have  $N$  training pixels  $\mathbf{x}_1, \dots, \mathbf{x}_N$ . In surveillance applications when the target class is rare or sparsely populated, the training pixels are usually taken as those surrounding the test pixel and assumed target-free [4]. Again similar to [4], we assume that  $\mathbf{x}_1, \dots, \mathbf{x}_N$  are independent and identically distributed (i.i.d.) Gaussian random vectors with zero mean and covariance matrix  $\mathbf{R}_b$ , and independent of the test pixel  $\mathbf{x}$ .

The problem in question is to find an efficient decision rule for the composite hypothesis testing problem (8.1) or (8.2), given knowledge of the test pixel  $\mathbf{x}$ , target signal signature  $\mathbf{s}$ , and training pixels  $\mathbf{x}_1, \dots, \mathbf{x}_N$ . Our goal is to achieve good detection performance for small  $N$ .

Before closing this section, we remark that our parametric detection schemes, as well as many others (e.g., in [4]), rely on the perfect knowledge of the target spectral signature. Generally, the target signature is available in its *reflectance spectrum*, whereas the HSI sensors measure the *radiance spectrum* of the observed materials. In order to apply these detection techniques, the

---

<sup>3</sup>The spectral signature may vary due to variations in atmospheric conditions and other factors, and the uncertainty can be captured by a linear mixing model [4]. We do not consider such spectral variations since our focus is effective cancellation of the background.



HSI data must be pre-processed to obtain reflectance data from the radiance ones (e.g., through atmospheric correction) or, alternatively, target spectral reflectance must be processed to obtain the radiance spectrum. See [70, 71] for details.

### 8.3 Covariance-Matrix Based Solutions

A number of solutions to the above problem have been developed. If the covariance matrix  $\mathbf{R}_b$  is known exactly, the optimum detector for (8.1) with unknown signal amplitude is the matched filter (MF) [12]:

$$\frac{|s^T \mathbf{R}_b^{-1} \mathbf{x}|^2}{s^T \mathbf{R}_b^{-1} s} \underset{H_0}{\overset{H_1}{\geq}} t_{\text{MF}}, \quad (8.3)$$

where  $t_{\text{MF}}$  denotes the MF threshold. The MF detector is obtained by a GLR approach (e.g., [7]), by which the ML estimate of the unknown amplitude  $a$  is first estimated and then substituted back into the likelihood ratio to form a test statistic. In practice, the MF detector *cannot* be implemented since  $\mathbf{R}_b$  is typically unknown. However, it provides a baseline for performance comparison when considering any realizable detection scheme.

In practice, the unknown  $\mathbf{R}_b$  can be replaced by some estimate, such as the sample covariance matrix obtained from the training pixels:

$$\hat{\mathbf{R}}_b = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n \mathbf{x}_n^T. \quad (8.4)$$

Using  $\hat{\mathbf{R}}_b$  in (8.3) leads to the so-called AMF detector [12]:

$$\frac{|s^T \hat{\mathbf{R}}_b^{-1} \mathbf{x}|^2}{s^T \hat{\mathbf{R}}_b^{-1} s} \underset{H_0}{\overset{H_1}{\geq}} t_{\text{AMF}}, \quad (8.5)$$

where  $t_{\text{AMF}}$  denotes the AMF threshold.

Alternatively, one can treat both  $a$  and  $\mathbf{R}_b$  as unknowns and estimate them successively by ML. Such a GLR approach was pursued by Kelly [13], which gives the following Kelly test:

$$\frac{|s^T \hat{\mathbf{R}}_b^{-1} \mathbf{x}|^2}{\left(s^T \hat{\mathbf{R}}_b^{-1} s\right) \left(N + \mathbf{x}^T \hat{\mathbf{R}}_b^{-1} \mathbf{x}\right)} \underset{H_0}{\overset{H_1}{\geq}} t_{\text{Kelly}}, \quad (8.6)$$

where  $t_{\text{Kelly}}$  denotes the corresponding threshold.

Another popular detector is the ACE test [14, 15]:

$$\frac{|s^T \hat{\mathbf{R}}_b^{-1} \mathbf{x}|^2}{\left(s^T \hat{\mathbf{R}}_b^{-1} s\right) \left(\mathbf{x}^T \hat{\mathbf{R}}_b^{-1} \mathbf{x}\right)} \underset{H_0}{\overset{H_1}{\geq}} t_{\text{ACE}}, \quad (8.7)$$

which is obtained by a GLR procedure that takes into account not only the unknown amplitude  $a$  and background covariance matrix  $\mathbf{R}_b$ , but also the variability of the variance of the background under  $H_0$  and  $H_1$ . Interestingly, the ACE test is the AMF test (8.5) normalized by the signal energy weighted by the covariance matrix inverse  $\hat{\mathbf{R}}_b^{-1}$ . By the Schwartz inequality, one can see that the ACE test statistic is bounded between zero and one.

The AMF, Kelly and ACE tests have constant false alarm rate (CFAR). However, they entail a large training requirement. The covariance matrix  $\mathbf{R}_b$  has a dimension of  $L \times L$ . Typical values for  $L$  in real HSI systems are in the range of hundreds. An accurate estimate of the covariance matrix would require a large number of target-free training pixels, which may not be available, especially in non-homogeneous environments. In addition, the computational complexity of these detectors is high, since  $\mathbf{R}_b$  has to be estimated and inverted frequently.

## 8.4 Proposed Approach

In this section, we present a class of parametric adaptive signal detectors with reduced training requirement. The proposed detectors, which are detailed in Section 8.4.2, relies on an NS-AR model introduced in Section 8.4.1, an ML parameter estimation algorithm derived in Section 8.4.3, a model order selection method discussed in Section 8.4.4, and a training screening technique presented in Section 8.4.5.

### 8.4.1 Parametric Modeling of HSI Data

It is well-known that the interference suppression ability of the detectors discussed in Section 8.3 comes from a whitening procedure. Consider, for example, the AMF detector (8.5). The whitening operation takes as inputs the signature vector  $\mathbf{s}$  and test pixel  $\mathbf{x}$ , and outputs whitened versions:

$$\tilde{\mathbf{s}} = \hat{\mathbf{R}}_b^{-1/2} \mathbf{s}, \quad \tilde{\mathbf{x}} = \hat{\mathbf{R}}_b^{-1/2} \mathbf{x}, \quad (8.8)$$

where  $\hat{\mathbf{R}}_b^{-1/2}$  denotes the matrix square-root of  $\hat{\mathbf{R}}_b^{-1}$ . Following the whitening, the AMF detector reduces to simple correlation of the whitened outputs:

$$\frac{|\tilde{\mathbf{s}}^T \tilde{\mathbf{x}}|^2}{\tilde{\mathbf{s}}^T \tilde{\mathbf{s}}} \underset{H_0}{\overset{H_1}{\gtrless}} t_{\text{MF}}. \quad (8.9)$$

If the whitening operation can be designed or approximated via a parametric model without explicitly estimating  $\mathbf{R}_b$ , then it is conceivable that fewer training pixels are needed, provided that the parametric model is parsimonious enough (without an extraordinary number of parameters). This is the essence of our parametric model based methods. Next, we consider two different parametric modeling approaches.

## AR Modeling

AR models have been popular choices for parametric modeling in spectral analysis, speech coding, wireless channel modeling, seismic signal processing, among others (e.g., [30]). Parametric adaptive detection based on multichannel AR models has been considered in [22]- [23, 72, 73] for airborne radar systems equipped with multiple antennas. It was shown that significant saving in training and complexity can be achieved by fitting the interference and radar clutter into suitable multichannel AR models.

For the problem under study, the  $L \times 1$  background vector  $\mathbf{b}$ , or equivalently the observed signal  $\mathbf{x}$  under  $H_0$ , may be assumed to be a scalar AR process which produces the  $L$  samples of  $\mathbf{b}$ . If an AR model is appropriate for HSI data, then the detection problem amounts to first estimating the AR coefficients from training data, whitening the signals by a whitening filter constructed from the AR coefficient estimates, and computing the decision statistic from the whitened signals followed by thresholding. For brevity, the above approach is referred to as the *parametric adaptive matched filter (PAMF)*<sup>4</sup>, or *normalized PAMF (NPAMF)* [72] if the decision variable is normalized, similar to the normalization imposed by the ACE detector of (8.7).

We have tested the above AR-based PAMF/NPAMF detectors with real HSI data using fixed AR parameters across the spectral domain and found they suffer a performance loss compared to the methods proposed here. The reason is that AR models are not a suitable parametric model for HSI data. In particular, we find that HSI data are *non-stationary in the spectral dimension*, whereas fixed parameter AR models characterize *stationary* random processes. To see this, we have computed the sample covariance matrix  $\hat{\mathbf{R}}_b$ , from a total of  $K = 24 \times 46 = 1104$  training pixels drawn from a homogeneous region of the HSI data described in Section 8.5. Figure 8.1 depicts the main and 3 sub-diagonals of  $\hat{\mathbf{R}}_b$ , which correspond to the autocorrelation function (ACF) at spectral lag 0 (i.e., variance), lag 1, lag 2 and lag 3, respectively, versus the spectral bands. Clearly, the signal is not stationary since the variance and ACF at other lags vary significantly across the spectral bands.

## NS-AR Modeling

Although HSI data is non-stationary (NS) across the entire spectral dimension, it may be considered approximately stationary over a sufficiently small number of adjacent spectral bands. This can be seen from Figure 8.1 where the variation of the sample statistics over a few adjacent spectral bands is considerably smaller compared with that over the entire spectral bands. In the following, we consider a NS-AR modeling approach by taking into account such *local stationarity* of HSI data. Specifically, let  $x_n(l)$  denote the spectral response at the  $l$ th spectral band of the  $n$ th training pixel  $\mathbf{x}_n$ , that is,  $\mathbf{x}_n = [x_n(0), \dots, x_n(L-1)]^T$ . Then, we slice  $\mathbf{x}_n$  into  $L - L_s + 1$  overlapping subvectors:

$$\begin{aligned} \mathbf{x}_{n,l} &= [x_n(l), \dots, x_n(l + L_s - 1)]^T, \\ l &= 0, \dots, L - L_s, \end{aligned} \quad (8.10)$$

---

<sup>4</sup>Details of the PAMF and NPAMF detectors can be inferred from the proposed NS-PAMF and NS-NPAMF detectors discussed in Section 8.4.2, as the former are special cases of the latter.

where  $L_s \leq L$  denotes the length of the subvectors. Equivalently, these subvectors can be thought of as being obtained by windowing  $\mathbf{x}_n$  using a sliding window of size  $L_s$ . For sufficiently small  $L_s$ , each subvector  $\mathbf{x}_{n,l}$  can be modeled as an  $M$ th-order AR process:

$$\begin{aligned} x_n(k) &= - \sum_{m=1}^M a_l(m) x_n(k-m) + w_{n,l}(k), \\ k &= l, l+1, \dots, l+L_s-1; \quad n = 1, \dots, N, \end{aligned} \quad (8.11)$$

where  $w_{n,l}(k)$  denotes the modeling residual for the  $l$ th subvector  $\mathbf{x}_{n,l}$ . The residual is Gaussian (since  $x_n(k)$  is so) with zero-mean and variance  $\sigma_l^2$ , and spectrally white so that  $\{w_{n,l}(k)\}$  are independent with respect to  $k$  and  $n$  [30]. Note that the  $l$ th set of the AR coefficients,  $a_l(1), \dots, a_l(M)$ , is associated with the  $l$ th subvector  $\mathbf{x}_{n,l}$ , and that different subvectors are associated with different sets of AR coefficients. For simplicity, we consider fixed AR model order  $M$  (also see discussions in Section 8.4.4).

From the estimation perspective, the choice of  $M$  and window size  $L_s$  should be made with tradeoffs among the bias, variance and stationarity of the modeling approach. A large  $M$  might be desirable since it can provide better fitting (lower bias) to the HSI data. Increasing  $M$ , however, would require the window size  $L_s$  to increase accordingly since more parameters are to be estimated and, therefore, more data should be provided within each subvector to reduce the variance of parameter estimates. If  $L_s$  is too large, the assumption of stationarity within the subvector may be violated, which can cause significant degradation. From the application aspect, however, these parameters are related to the HSI sensor characteristics, such as the operating spectral range, spectral resolution, etc. For the HSI data used in this chapter, we found a window size  $8 \leq L_s \leq 15$  is generally appropriate for modeling. Once  $L_s$  is selected, we can use information criterion based model order selection techniques to determine  $M$ . We leave the details to Section 8.4.4.

Instead of the above sliding-window based NS-AR modeling approach, one can consider an alternative NS-AR model that models the HSI data across all the spectral bands:

$$\begin{aligned} x_n(l) &= - \sum_{m=1}^M b_l(m) x_n(l-m) + v_n(l), \\ l &= 0, \dots, L-1, \end{aligned} \quad (8.12)$$

where  $b_l(m)$  denotes the shift-varying AR coefficient and  $v_n(l)$  the fitting error of the  $l$ th sample. Note that the above model differs from (8.11) in that the AR coefficients are varying from sample to sample, whereas in (8.11), they are assumed to remain fixed within a sliding subvector of  $L_s$  samples. An additional parametric model for the shift-varying AR coefficients  $\{b_l(m)\}$  is necessary to ensure they can be estimated. This *doubly* parametric approach is more sensitive to the choice of the parameters, whose estimation is also considerably more involved. In the following, we consider only the sliding-window based NS-AR modeling approach.

## 8.4.2 NS-AR Model Based Parametric Adaptive Detectors

If the above NS-AR model (8.11) is appropriate for modeling target-free HSI data (i.e., the background), then a time-series based (as opposed to the previous covariance-matrix based) whitening process can be developed without explicitly estimating  $\mathbf{R}_b$ . This leads to a class of parametric adaptive detectors that are summarized below:

- **Step 1 – Parameter Estimation:** Estimate the NS-AR coefficients  $\{a_l(m)\}$  in (8.11) and the variance  $\{\sigma_l^2\}$  of the residual from the training pixels  $\{\mathbf{x}_n\}_{n=1}^N$  by using an ML based estimation algorithm detailed in Section 8.4.3. Let  $\{\hat{a}_l(m), \hat{\sigma}_l^2\}$  denote the coefficient estimates.
- **Step 2 – Whitening:** Form a shift-varying moving-average (MA) whitening filter from the parameter estimates  $\{\hat{a}_l(m), \hat{\sigma}_l^2\}$ , and whiten the test pixel  $\mathbf{x}$  and target signature  $\mathbf{s}$  as follows:

$$\begin{aligned}\check{x}(l) &= \frac{1}{\hat{\sigma}_l} \left[ x(l) + \sum_{m=1}^M \hat{a}_{l-L_s}(m) x(l-m) \right], \\ \check{s}(l) &= \frac{1}{\hat{\sigma}_l} \left[ s(l) + \sum_{m=1}^M \hat{a}_{l-L_s}(m) s(l-m) \right], \\ l &= L_s - 1, \dots, L - 1,\end{aligned}\tag{8.13}$$

where  $\check{x}(l)$  and  $\check{s}(l)$  denote the  $l$ th output sample of the whitening filter when the input is the test pixel  $\mathbf{x}$  and target signature  $\mathbf{s}$ , respectively. It should be noted from (8.13) that each set of the NS-AR parameter estimates, i.e.,  $\{\hat{a}_l(m)\}_{m=1}^M$  and  $\hat{\sigma}_l$ , is used to compute one pair of output samples  $\check{x}(l)$  and  $\check{s}(l)$ ; as the sliding window shifts to the next position, we use the next set of parameter estimates for whitening. In effect, (8.13) implements the whitening operation (8.8) in a time-series fashion by taking into account the NS nature of the signal. For an input of  $L$  spectral samples, the time-series based whitening filter outputs  $L - L_s$  whitened samples due to initialization of the whitening filter. Although such dimensionality reduction may affect the detection performance, the impact is negligible for small  $L_s$  and large  $L$ , which is typical in HSI systems.

- **Step 3 – Detection:** The outputs of the shift-varying whitening filter corresponding to the test pixel  $\mathbf{x}$  and target signature  $\mathbf{s}$ , respectively, are used to form the decision statistic. Depending upon how the decision statistic is formed, we have a class of parametric detectors. For example, the parametric counterparts of the covariance-matrix based AMF (8.5) and

ACE (8.7) detectors are given by

$$\frac{\left| \sum_{l=L_s-1}^{L-1} \check{s}(l)\check{x}(l) \right|^2}{\sum_{l=L_s-1}^{L-1} \check{s}^2(l)} \underset{H_0}{\overset{H_1}{\gtrless}} t_{\text{NS-PAMF}}, \quad (8.14)$$

$$\frac{\left| \sum_{l=L_s-1}^{L-1} \check{s}(l)\check{x}(l) \right|^2}{\left( \sum_{l=L_s-1}^{L-1} \check{s}^2(l) \right) \left( \sum_{l=L_s-1}^{L-1} \check{x}^2(l) \right)} \underset{H_0}{\overset{H_1}{\gtrless}} t_{\text{NS-NPAMF}}, \quad (8.15)$$

which are referred to as the *NS-PAMF* and *NS-NPAMF* detectors, respectively. A time-series based, Kelly like test can also be obtained in a similar fashion.

Finally, it is noted that the above NS-PAMF and NS-NPAMF detectors reduce to the PAMF and NPAMF detectors, respectively, that are briefly discussed in Section 8.4.1, when  $L_s = L$ , that is, the sliding window reaches the maximum value and includes the entire spectral bands. In that case, the NS-AR model in (8.11) reduces to the standard stationary AR model.

### 8.4.3 ML Estimation of NS-AR Coefficients

Parameter estimation plays a critical role for the proposed parametric detectors. In this section, we present an ML estimator to estimate the NS-AR coefficients in (8.11) using training pixels  $\mathbf{x}_1, \dots, \mathbf{x}_N$ . Our ML estimator is an extension of that in [30] for fixed AR models to NS-AR processes.

Consider the vector of AR coefficients of model order  $M$ :  $\mathbf{a}_l = [a_l(1), \dots, a_l(M)]^T$ . According to the statistical assumptions made in Section 8.2 and the NS-AR model (8.11), the  $l$ th set of subvectors  $\mathbf{x}_{1,l}, \dots, \mathbf{x}_{N,l}$  formed from the  $N$  training pixels are i.i.d. multivariate Gaussian whose joint probability density function (PDF) is parameterized by the NS-AR coefficients  $\mathbf{a}_l$  and variance  $\sigma_l^2$ . Then, the ML estimates of  $\mathbf{a}_l$  and  $\sigma_l^2$  are obtained by maximizing the joint PDF  $p(\mathbf{x}_{1,l}, \dots, \mathbf{x}_{N,l}; \mathbf{a}_l, \sigma_l^2)$ . Exact maximization of the joint PDF with respect to the unknown parameters turns out to be highly involved computationally [74]. Instead, we seek to optimize a conditional PDF, which produces an asymptotic ML estimate of the parameters for large data size [30]. Specifically, let

$$\mathbf{x}_{n,l}^{(1)} = [x_n(l), \dots, x_n(l+M-1)]^T, \quad (8.16)$$

$$\mathbf{x}_{n,l}^{(2)} = [x_n(l+M), \dots, x_n(l+L_s-1)]^T \quad (8.17)$$

which collect the first  $M$  and, respectively, the last  $L_s - M$  samples of  $\mathbf{x}_{n,l}$ . Thus, we have  $\mathbf{x}_{n,l} = [\mathbf{x}_{n,l}^{(1)T}, \mathbf{x}_{n,l}^{(2)T}]^T$ . Our asymptotic ML estimator seeks to maximize the joint conditional PDF

$p\left(\mathbf{x}_{1,l}^{(2)}, \dots, \mathbf{x}_{N,l}^{(2)} \middle| \mathbf{x}_{1,l}^{(1)}, \dots, \mathbf{x}_{N,l}^{(1)}; \mathbf{a}_l, \sigma_l^2\right)$  with respect to  $\mathbf{a}_l$  and  $\sigma_l^2$ . We will write the conditional PDF as  $p\left(\mathbf{x}_l^{(2)} \middle| \mathbf{x}_l^{(1)}; \mathbf{a}_l, \sigma_l^2\right)$  for brevity.

To find an explicit form of the above conditional PDF, we observe from (8.11) that

$$w_{n,l}(k) = x_n(k) + \sum_{m=1}^M a_l(m)x_n(k-m), \quad (8.18)$$

$$k = l, l+1, \dots, l+L_s-1; n = 1, \dots, N.$$

Since  $\{w_{n,l}\}$  are i.i.d. Gaussian with zero mean and variance  $\sigma_l^2$ , we have (e.g., [30])

$$\begin{aligned} & p\left(\mathbf{x}_l^{(2)} \middle| \mathbf{x}_l^{(1)}; \mathbf{a}_l, \sigma_l^2\right) \\ &= (2\pi\sigma_l^2)^{-N(L_s-M)/2} \exp\left\{-\frac{1}{2\sigma_l^2} \sum_{n=1}^N \right. \\ & \quad \left. \cdot \sum_{k=l+M}^{l+L_s-1} \left[x_n(k) + \sum_{m=1}^M a_{m,l}x_n(k-m)\right]^2\right\} \end{aligned} \quad (8.19)$$

Maximizing the above conditional PDF is equivalent to minimizing the negative log likelihood function

$$V(\mathbf{a}_l, \sigma_l^2) = -\ln p\left(\mathbf{x}_l^{(2)} \middle| \mathbf{x}_l^{(1)}; \mathbf{a}_l, \sigma_l^2\right). \quad (8.20)$$

Define an  $(L_s - M) \times M$  matrix

$$\mathbf{X}_{n,l} = \begin{bmatrix} x_n(l+M-1) & \dots & x_n(l) \\ \vdots & \vdots & \vdots \\ x_n(l+L_s-2) & \dots & x_n(l+L_s-M-1) \end{bmatrix}. \quad (8.21)$$

Then,  $V(\mathbf{a}_l, \sigma_l^2)$  can be more compactly expressed as

$$\begin{aligned} V(\mathbf{a}_l, \sigma_l^2) &= C_1 + \frac{1}{2}N(L_s - M) \ln \sigma_l^2 \\ & \quad + \frac{1}{2\sigma_l^2} \sum_{n=1}^N \left\| \mathbf{x}_{n,l}^{(2)} + \mathbf{X}_{n,l} \mathbf{a}_l \right\|^2, \end{aligned} \quad (8.22)$$

where<sup>5</sup>

$$C_1 = \frac{1}{2}N(L_s - M) \ln(2\pi), \quad (8.23)$$

Taking the derivative of  $V(\mathbf{a}_l, \sigma_l^2)$  with respect to  $\sigma_l^2$  and setting it to zero yield

$$\hat{\sigma}_l^2(\mathbf{a}_l) = \frac{1}{N(L_s - M)} \sum_{n=1}^N \left\| \mathbf{x}_{n,l}^{(2)} + \mathbf{X}_{n,l} \mathbf{a}_l \right\|^2. \quad (8.24)$$

---

<sup>5</sup>We keep the constant term  $C_1$  which depends on  $M$  for model order selection in Section 8.4.4.

Substituting  $\hat{\sigma}_l^2(\mathbf{a}_l)$  back into (8.22) reduces the cost function to

$$V(\mathbf{a}_l, \hat{\sigma}_l^2) = C_1 + C_2 + \frac{1}{2}N(L_s - M) \ln \hat{\sigma}_l^2(\mathbf{a}_l), \quad (8.25)$$

where

$$C_2 = \frac{1}{2}N(L_s - M). \quad (8.26)$$

Therefore, the ML estimate of  $\mathbf{a}_l$  is obtained by minimizing  $\hat{\sigma}_l^2(\mathbf{a}_l)$ , the variance of the NS-AR modeling residual. The solution is obtained by least-squares fitting:

$$\hat{\mathbf{a}}_l = - \left( \sum_{n=1}^N \mathbf{X}_{n,l}^T \mathbf{X}_{n,l} \right)^{-1} \left( \sum_{n=1}^N \mathbf{X}_{n,l}^T \mathbf{x}_{n,l}^{(2)} \right), \quad (8.27)$$

$$l = 0, 1, \dots, L - L_s.$$

The matrix within the first pair of brackets is assumed non-singular. A necessary condition for non-singularity is that the number of training pixels  $N$  is such that

$$N \geq \frac{M}{L_s - M}. \quad (8.28)$$

This is because the above matrix inverse can be expressed as  $(\mathbf{X}_l^T \mathbf{X}_l)^{-1}$ , where

$$\mathbf{X}_l = [\mathbf{X}_{1,l}^T, \dots, \mathbf{X}_{N,l}^T]^T \quad (8.29)$$

is a tall matrix when the above condition is satisfied. On the other hand, when  $N \geq \frac{M}{L_s - M}$ ,  $\mathbf{X}_l^T \mathbf{X}_l$  is full rank almost surely due to the random nature of the HSI data.

Finally, substituting the ML estimate (8.27) back into (8.24) yields the minimum variance of the residual:

$$\hat{\sigma}_l^2 = \frac{1}{N(L_s - M)} \mathbf{x}_l^{(2)T} \mathbf{P}_{\mathbf{X}_l}^\perp \mathbf{x}_l^{(2)}. \quad (8.30)$$

where  $\mathbf{x}_l^{(2)} = [\mathbf{x}_{1,l}^{(2)T}, \dots, \mathbf{x}_{N,l}^{(2)T}]$  and  $\mathbf{P}_{\mathbf{X}_l}^\perp$  is the projection matrix onto the null space of  $\mathbf{X}_l$ :

$$\mathbf{P}_{\mathbf{X}_l}^\perp = \mathbf{I} - \mathbf{X}_l (\mathbf{X}_l^T \mathbf{X}_l)^{-1} \mathbf{X}_l^T, \quad (8.31)$$

where  $\mathbf{I}$  is an identity matrix.

#### 8.4.4 NS-AR Model Order Selection

In this section, we develop information criterion based model order selection techniques to determine the NS-AR model order  $M$  in (8.11). Although in principle it is possible to select a different  $M$  for each subvector  $\mathbf{x}_{n,l}$ ,  $l = 0, \dots, L - L_s$ , by a separate fitting of  $M$  to the information criterion, this is a tedious process. In the following, we use a fixed  $M$  for all  $l$ .



Model order selection for parametric models is a classical topic and has been investigated by various researchers for various models (e.g., [30, 31] and references therein). We examine herein the model order selection problem for the NS-AR model in (8.11) for the HSI application, which appears not to have been addressed elsewhere. Specifically, we consider a *generalized Akaike information criterion (GAIC)*, which chooses the model order  $M$  that minimizes

$$W(M) = \sum_{l=0}^{L-L_s} [V_l(M) + \gamma(M)], \quad (8.32)$$

where  $V_l(M)$  is the minimum cost associated with the  $l$ th set of subvectors  $\mathbf{x}_{1,l}, \dots, \mathbf{x}_{N,l}$ , and  $\gamma(M)$  is a penalty term that penalizes increasing model order [31]. Specifically, the minimum cost is derived in Section 8.4.3 (cf. (8.25))

$$V_l(M) = C_1(M) + C_2(M) + \frac{1}{2}N(L_s - M) \ln \hat{\sigma}_l^2(M), \quad (8.33)$$

where  $C_1(M)$ ,  $C_2(M)$  and  $\hat{\sigma}_l^2(M)$  are given by (8.23), (8.26) and (8.30), respectively, and the dependence on  $M$  is made explicit. On the other hand, the penalty term typically takes the form [31]

$$\gamma(M) = \alpha(M + 1) \ln(NL_s), \quad (8.34)$$

or

$$\gamma(M) = \alpha(M + 1) \ln [\ln(NL_s)], \quad (8.35)$$

where  $M + 1$  is the total number of unknowns for each set of subvectors  $\{\mathbf{x}_{n,l}\}_{n=1}^N$ ,  $NL_s$  is the number of data samples contained in  $\{\mathbf{x}_{n,l}\}_{n=1}^N$ , and  $\alpha \geq 2$  is a parameter of user choice. Note that the above GAIC reduces to the standard AIC [75] when the  $(L - L_s + 1)$ -term summation in (8.32) vanishes and  $\gamma(M) = 2(M + 1)$ . It is known that AIC is not a consistent model order estimator [30]. Choosing a penalty term proportional to  $\ln(NL_s)$  or  $\ln [\ln(NL_s)]$  is an effective way of obtaining consistent order estimates [31].

### 8.4.5 Training Screening

One assumption made in Section 8.2 is that the  $N$  training pixels  $\mathbf{x}_1, \dots, \mathbf{x}_N$  are target-free. This assumption is reasonable in *homogeneous* environments where targets are rare or sparsely populated, but usually violated in *heterogeneous* or *dense-target* environments. In the latter case, the performance of all training-based detectors, including those covariance-matrix based detectors discussed in Section 8.3, degrade considerably. Training screening to eliminate “bad” training data in such cases has been examined in a number of recent studies for radar target detection (e.g., [76]- [78] and references therein). In this section, we discuss screening of heterogeneous HSI training data. Rather than treating it as an independent process, we cast training screening within the proposed NS-AR framework.

For covariance-matrix based detectors in Section 8.3, one screening approach according to statistical ranking and selection theory is to compute the following metric from the training set [76]:

$$T_n = \mathbf{x}_n^T \hat{\mathbf{R}}_b^{-1} \mathbf{x}_n, \quad n = 1, \dots, N. \quad (8.36)$$

Then, the metric is used to partition the training set  $\mathcal{S} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  into two disjoint sets  $\mathcal{S}_1$  and  $\mathcal{S}_2$  (see [76] for details), of which the former contains the refined training data while the latter contains *outliers* that are discarded.

The above training screening approach relies on an estimate of a full-rank sample covariance matrix  $\hat{\mathbf{R}}_b$ . To circumvent this, we note that  $\mathbf{x}_n^T \hat{\mathbf{R}}_b^{-1} \mathbf{x}_n = \|\tilde{\mathbf{x}}_n\|^2$ , where  $\tilde{\mathbf{x}} = \hat{\mathbf{R}}_b^{-1/2} \mathbf{x}_n$ , i.e., the “whitened” version of  $\mathbf{x}_n$ . The whitening operation can be equivalently implemented in a time-series fashion by an MA whitening filter without the need to estimate  $\hat{\mathbf{R}}_b$ . This alternative screening approach is proposed in [78] and referred to as the *innovation power sorting (IPS)* method, since the output of the MA whitening filter is often called the *innovation* of the input (e.g., [79]).

The IPS can be extended and cast within the NS-AR framework. Specifically, we first use the ML estimator in Section 8.4.3 to estimate the NS-AR parameters  $\{\hat{a}_l(m), \hat{\sigma}_l^2\}$  from the original training set  $\mathcal{S}$ . Next, we form a shift-varying MA whitening filter from these parameter estimates and, similarly to (8.13), whiten the training set as follows:

$$\check{x}_n(l) = \frac{1}{\hat{\sigma}_l} \left[ x_n(l) + \sum_{m=1}^M \hat{a}_{l-L_s}(m) x_n(l-m) \right], \quad (8.37)$$

$$l = L_s - 1, \dots, L - 1; n = 1, \dots, N.$$

Finally, we compute the following metric

$$T_n = \sum_{l=L_s-1}^{L-1} \check{x}_n^2(l), \quad n = 1, \dots, N, \quad (8.38)$$

which is used to replace (8.36) for the partition of  $\mathcal{S}$  into  $\mathcal{S}_1$  and  $\mathcal{S}_2$ .

## 8.5 Experimental Results

In this section, we present experimental results to illustrate the performance of our proposed techniques. For comparison, we consider the covariance matrix based **ACE** test (8.7), the AR model based **NPAMF** detector (see Section 8.4.1), our NS-AR model based **NS-NPAMF** detector (8.15), and a modified version called **NS-LP-NPAMF** that is briefly explained below. We do not compare with the AMF (8.5) or Kelly (8.6) tests which were found to perform similarly to the ACE test in our experiments. Meanwhile, the ACE, NPAMF, NS-NPAMF and LS-NS-NPAMF are all normalized tests whose test statistics range between 0 and 1, which makes comparison more convenient.

The modification made in the NS-LP-NPAMF detector is due to an observation that HSI spectral data exhibit small oscillations. As an example, Figure 8.2 depicts the original HSI data of a randomly chosen pixel from the HSI data set described below. Such oscillations along the spectral dimension do not contribute much to detection, meanwhile making parameter estimates more noisy. It was found that passing the HSI data through a lowpass (LP) filter to first remove those

oscillations before applying the proposed NS-AR modeling, estimation, and detection techniques is helpful. Our NS-NPAMF detector (8.15) with such a modification is called *NS-LP-NPAMF*. For lowpass filtering, we use a simple moving-average filter with impulse response given by a Kaiser window, whose length is equal to the sliding window size  $L_s$  and the shape parameter is 3. It should be noted that LP filtering is applied to all signals involved in detection, including the training pixels, test pixel, and target signature.

The HSI data employed in our studies is provided on the CD that accompanies [60]. Figure 8.3 is a color infrared (IR) image from a portion of the data set, which shows a view of an airborne hyperspectral data flightline over the Washington DC area. The sensor system used in this case measured the spectral response in 210 spectral bands in the 0.4 to 2.4  $\mu\text{m}$  region of the visible and IR spectrum. Bands in the 0.9 and 1.4  $\mu\text{m}$  region where the atmosphere is opaque have been omitted from the data, leaving  $L = 191$  spectral bands. Additional information on the data set can be found in [60]. The image shown in Figure 8.3 was made using bands 60, 27 and 17 for the red, green and blue colors, respectively. Three test regions are highlighted in Figure 8.3. Test region #1 is relatively homogeneous and formed by grass, test region #2 is less homogeneous with tree and road, and test region #3 corresponds to a heterogeneous environment. To simulate the  $H_1$  condition, we superimpose a target signal to the test pixel. The target signal corresponds to the spectral signature of a man-made object (taken from a pixel in Figure 8.3), and is scaled according to particular target fill factors [4]. Each test data set is first demeaned using a  $3 \times 3$  spatial moving average filter (see [80] for details on the demeaning process).

### 8.5.1 Model Order Selection

We first use the GAIC developed in Section 8.4.4 to determine the model order  $M$  of the NS-AR model. Figure 8.4 depicts  $W(M)$  in (8.32) as a function of  $M$  for  $L_s = 10$  and  $N = 8$ , and the result is obtained by averaging over the pixels in test region #1. Results obtained with the other two test regions are similar. It is seen that  $W(M)$  decreases quickly as  $M$  increases from 1 to 3, reaches its minimum and remains relatively flat between 3 and 5, then increases slightly from 5 to 8, and finally drops drastically for  $M = 9$ . The pattern of decrease followed by increase of  $W(M)$  is standard for most model selection techniques [30]. To understand why  $W(M)$  drops again at  $M = 9$ , we note that (8.28) is violated with  $M = 9$ . As a result,  $\mathbf{X}_l$  in (8.29) does not have full column rank, and there are numerous solutions for the NS-AR coefficients  $\{a_l(m)\}$  that lead to zero residual in the NS-AR model. In the following, we choose  $M = 5$ .

### 8.5.2 Detection in Homogeneous Environments

To illustrate detection performance in homogeneous environments, the figure of merit employed here is the *separation* of test statistics under  $H_0$  and  $H_1$ , which is also used in [4]. For all methods, we use  $N = 8$  training pixels, which corresponds to a  $3 \times 3$  region without counting the center pixel (i.e., test pixel), for sample covariance matrix or parameter estimation. The sample covariance matrix  $\hat{\mathbf{R}}_b$  is rank deficient in this case. As suggested in [4], we use the approximation  $\hat{\mathbf{R}}_b^{-1} \approx \mathbf{I} - \mathbf{U}_1 \mathbf{U}_1^T$ , where  $\mathbf{U}_1$  is formed by the principle eigenvectors of  $\hat{\mathbf{R}}_b$ , for

the ACE detector. The subvector length (i.e., sliding window length) is  $L_s = 10$  for NS-NPAMF and NS-LP-NPAMF.

First consider test region #1. Figures 8.5(a) to 8.5(d) depict the test statistic separation of the four detectors, respectively, as a function of the target fill factor. We note that NPAMF is the worst of all detectors, which corroborates our earlier observation that stationary AR modeling is not suitable for HSI data. However, both NS-NPAMF and NS-LP-NPAMF outperform the ACE test, with NS-LP-NPAMF being slightly better than NS-NPAMF. Specifically, we see that the former achieves full target-background separation when the fill factor is 0.25, while the latter does not.

Figures 8.6(a) to 8.6(d) depict the counterpart results when the detectors are applied to test region #2, which is less homogeneous than test region #1. It is seen that all four detectors experience some degradation relative to the previous results. However, the proposed NS-NPAMF and NS-LP-NPAMF detectors, especially the latter, still significantly outperform the others.

### 8.5.3 Detection in Heterogeneous Environments

We now consider detection in heterogeneous environments. To this end, we embed 5 targets at randomly chosen locations in test region #3. We run the ACE and NS-LP-NPAMF detectors throughout the test region pixel by pixel, with and without training screening. If training screening is not applied, we use the  $N = 8$  pixels surrounding the test pixel for training. Otherwise, we first compute metric (8.36) for the ACE detector and, respectively, metric (8.38) for the NS-LP-NPAMF detector using all pixels within the test region, and then the metrics are used to select  $N = 8$  new training pixels to refine the parameter/covariance matrix estimate. Figures 8.7(a) to 8.7(d) depict the test statistics of the two detectors, with and without training screening, versus the index of the pixels within the test region. The dotted lines in these plots indicate the indices/locations of the embedded targets. By comparing the results, it is seen that training screening helps both detectors. It is also seen that the proposed NS-LP-NPAMF detector outperforms the ACE detector with or without training screening.

Finally, we consider a dense-target scenario by embedding not only 5 targets but also outliers in test region #1. In particular, about 20% of the pixels at random locations in the region are embedded with outliers that have a different spectral signature from that of the target. Figures 8.8(a) to 8.8(d) show the test statistics of the ACE and NS-LP-NPAMF detectors with and without training screening. It is seen that the NS-LP-NPAMF detector overall achieves a better performance than the other.

## 8.6 Conclusions

In this chapter, we have exploited parametric modeling of HSI data and investigated its application for subpixel target detection in HSI systems. We have shown that HSI data are non-stationary in the spectral dimension, which makes parametric adaptive modeling and detection more challenging than earlier studies for stationary data. To deal with non-stationarity, we have proposed a

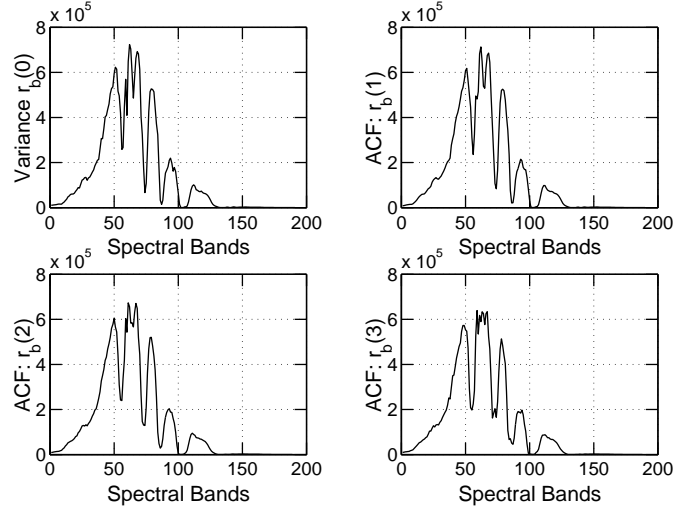


Figure 8.1: Sample estimates of the autocorrelation function (ACF) at spectral lag 0 (variance), lag 1, lag 2 and lag 3 across the spectral bands.

sliding-window based NS-AR model tailored for HSI data. We have developed parametric adaptive detectors by exploiting the NS-AR model, and addressed a range of issues including model order selection, training screening, parameter estimation, time-series based signal whitening, and detection. We have examined the performance of the proposed detectors and compared with covariance-matrix based techniques using real HSI data. It has been shown that the proposed parametric detectors are more efficient in training data usage and outperform the covariance-matrix based methods when training is limited.

Our approach implicitly assumes that HSI data is spectrally correlated. In most cases, HSI sensors oversample the spectral signal [3], which brings in spectral correlation in HSI data. The covariance-matrix based detectors, however, can be applied in the absence of spectral correlation (as in earlier multispectral systems with a few spectral bands [65]). While the covariance-matrix based AMF, ACE and GLRT detectors have a CFAR behavior, it is unclear whether the proposed detectors retain the same property. This remains an issue to be examined in the future. Other research along the proposed direction includes analytical study of the proposed detectors and exploration of alternative non-stationary parametric models for HSI target detection.

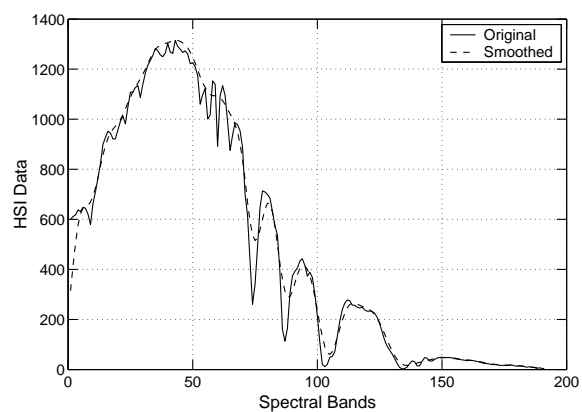


Figure 8.2: Original (solid line) and smoothed (dotted line) HSI data.

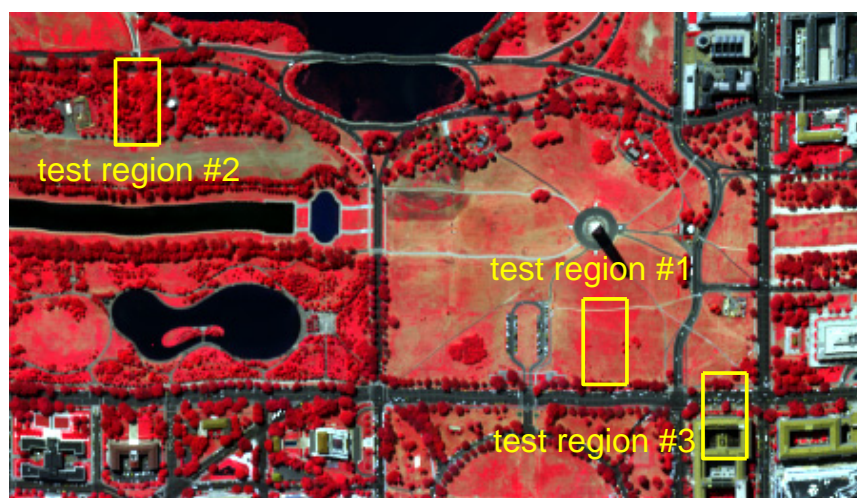


Figure 8.3: HSI image of the Washington DC Mall with  $L=191$  spectral bands. Three test regions are highlighted in yellow.

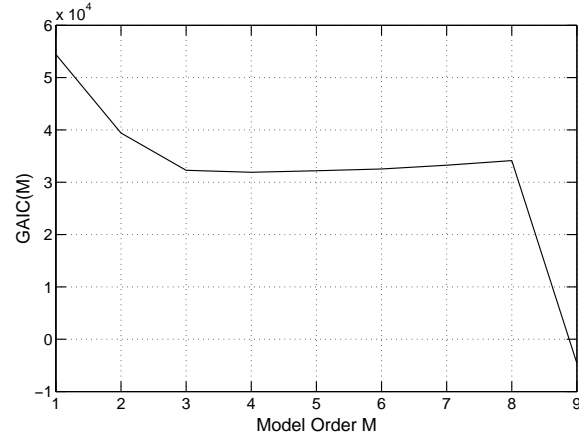


Figure 8.4: NS-AR model order selection.

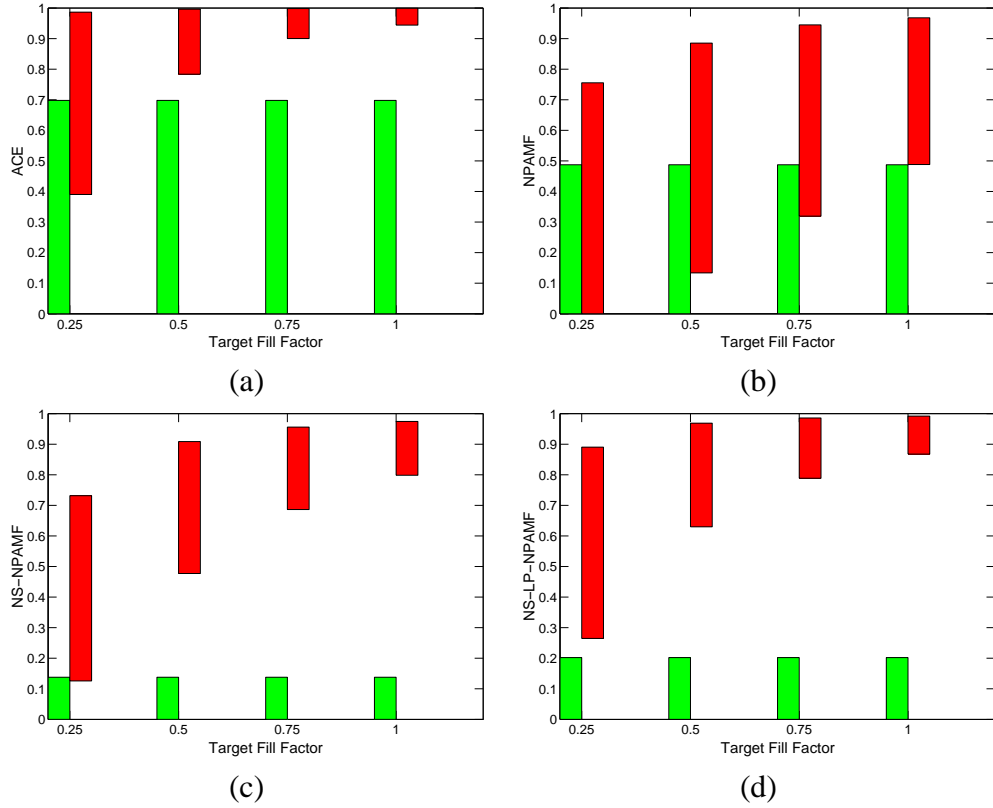


Figure 8.5: Test region #1: target-background separation versus target fill factor, where the red (dark) bars correspond to the range of test statistics under  $H_1$ , while the green (light) bars show the counterpart under  $H_0$ . (a) ACE. (b) NPAMF. (c) NS-NPAMF. (d) NS-LP-NPAMF.

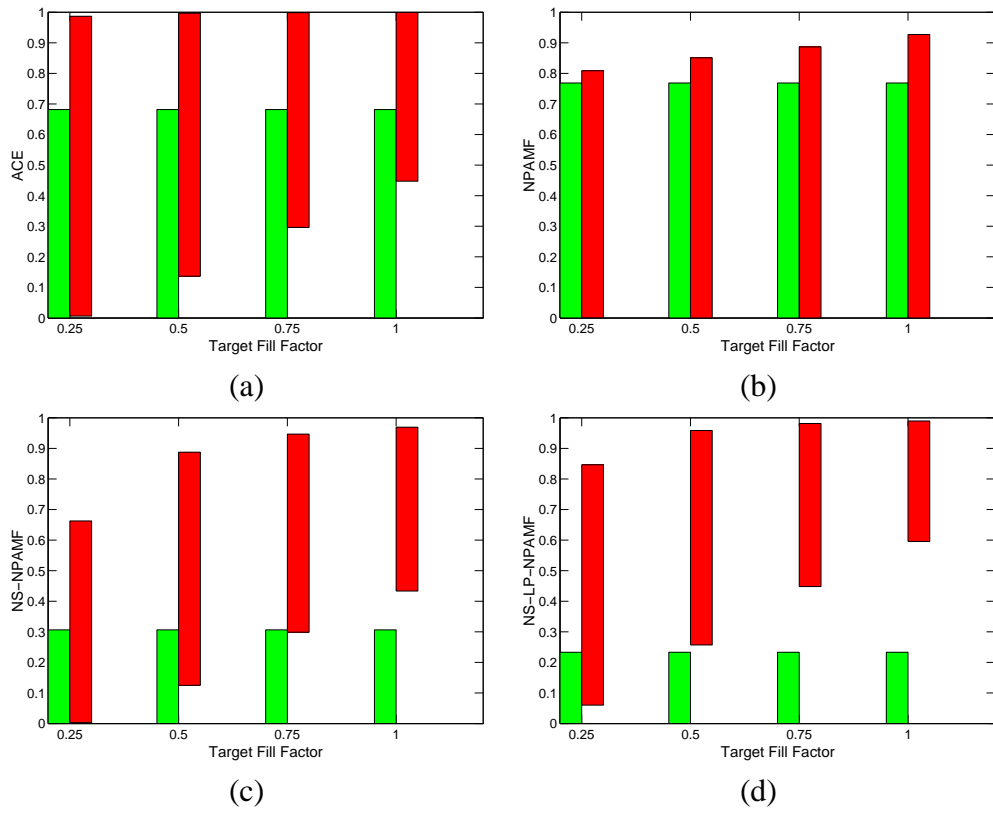


Figure 8.6: Test region #2: target-background separation versus target fill factor, where the red (dark) bars correspond to the range of test statistics under  $H_1$ , while the green (light) bars show the counterpart under  $H_0$ . (a) ACE. (b) NPAMF. (c) NS-NPAMF. (d) NS-LP-NPAMF.



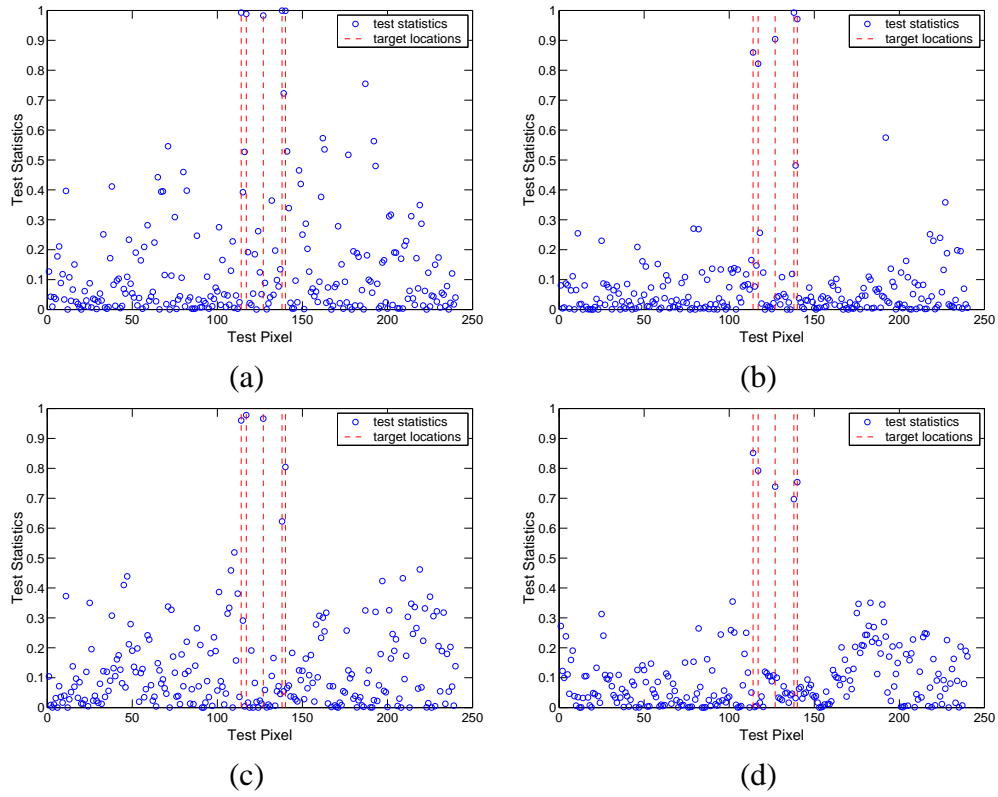


Figure 8.7: Test statistics of ACE and NS-LP-NPAMF of test pixels in the test region #3 with 5 embedded targets. (a) ACE without training screening. (b) NS-LP-NPAMF without training screening. (c) ACE with training screening. (d) NS-LP-NPAMF with training screening.

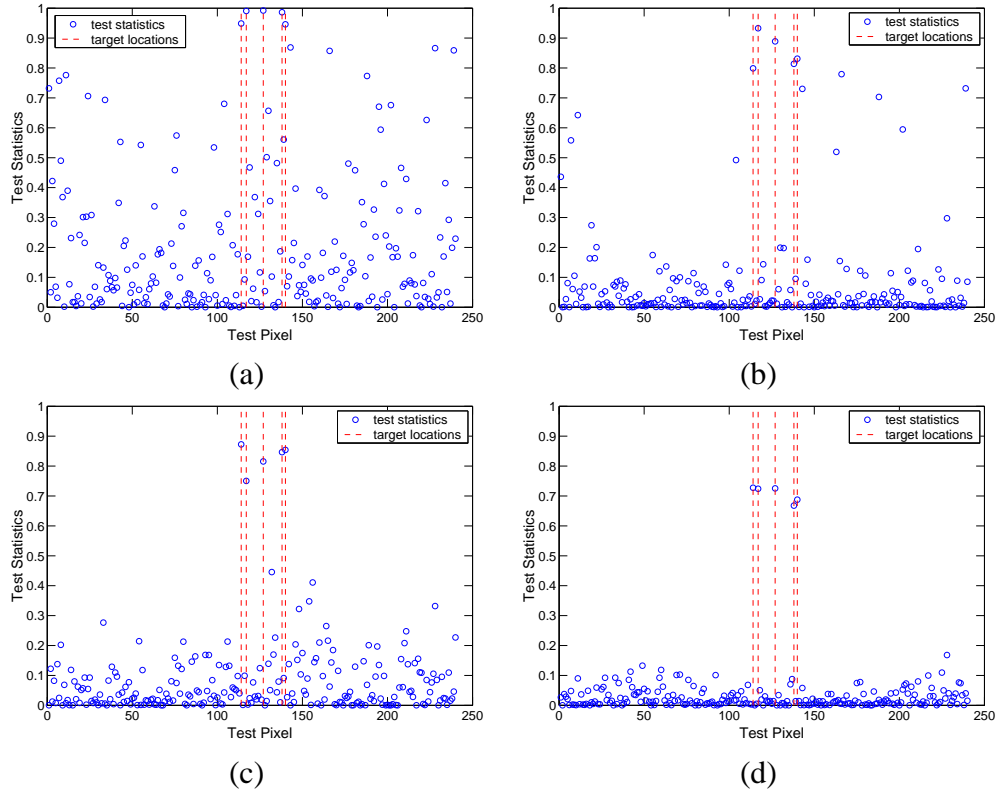


Figure 8.8: Test statistics of ACE and NS-LP-NPAMF of test pixels in the test region #1 with 5 embedded targets and more than 20% of the pixels are embedded with outliers. (a) ACE without training screening. (b) NS-LP-NPAMF without training screening. (c) ACE with training screening. (d) NS-LP-NPAMF with training screening.

# Appendix A

## Parametric Rao Test

### A.1 ML Parameter Estimation

In the following, we derive the ML estimates of the nuisance parameters  $\mathbf{Q}$  and  $\{\mathbf{A}(p)\}_{p=1}^P$  or  $\mathbf{A}$  defined in (3.15) under  $H_0$ , which will be needed in the derivation of the Rao test in Appendix A.2. The joint PDF or likelihood function  $\prod_k f_i(\mathbf{x}_k(0), \mathbf{x}_k(1), \dots, \mathbf{x}_k(N-1); \alpha, \mathbf{A}, \mathbf{Q})$  under  $H_i$ ,  $i = 0$  or  $1$ , can be written as

$$\begin{aligned} & \prod_k f_i(\mathbf{x}_k(P), \mathbf{x}_k(P+1), \dots, \mathbf{x}_k(N-1) | \mathbf{x}_k(0), \\ & \quad \mathbf{x}_k(1), \dots, \mathbf{x}_k(P-1); \mathbf{A}, \mathbf{Q}) \\ & \times f(\mathbf{x}_k(0), \mathbf{x}_k(1), \dots, \mathbf{x}_k(P-1); \alpha, \mathbf{A}, \mathbf{Q}). \end{aligned} \quad (\text{A.1})$$

The exact maximization of the PDF with respect to the unknown parameters produces a set of highly nonlinear equations that are difficult to solve. For large data records, the likelihood function can be approximated well by the conditional PDF in the above equation [35] and, therefore, the latter can be used for parameter estimation. After some manipulations using the standard procedure for obtaining the PDF of a set of transformed random variables, we have

$$\begin{aligned} & f_i(\mathbf{x}_k(P), \mathbf{x}_k(P+1), \dots, \mathbf{x}_k(N-1) | \mathbf{x}_k(0), \mathbf{x}_k(1), \\ & \quad \dots, \mathbf{x}_k(P-1); \alpha, \mathbf{A}, \mathbf{Q}) \\ & = \prod_{n=P}^{N-1} \frac{1}{\pi^J |\mathbf{Q}|} \exp \{ -\boldsymbol{\varepsilon}_k^H(n) \mathbf{Q}^{-1} \boldsymbol{\varepsilon}_k(n) \}. \end{aligned} \quad (\text{A.2})$$

where for  $k \geq 1$ ,  $\boldsymbol{\varepsilon}_k(n)$  is a function of the observed signals given by (7.3), whereas  $\boldsymbol{\varepsilon}_0(n)$  is given by (3.8) or (3.11) with  $\alpha = 0$  when  $i = 0$  and  $\alpha \neq 0$  when  $i = 1$ .

Recall that the training signals  $\{\mathbf{x}_k\}_{k=1}^K$  and the test signal  $\mathbf{x}_0$  are independent. Let  $\mathbf{X}(n) =$

$[\mathbf{x}_0^T(n), \mathbf{x}_1^T(n), \mathbf{x}_2^T(n), \dots, \mathbf{x}_K^T(n)]^T$ . The joint conditional PDF is given by

$$\begin{aligned} & f_i(\mathbf{X}(P), \mathbf{X}(P+1), \dots, \mathbf{X}(N-1) | \mathbf{X}(0), \mathbf{X}(1), \\ & \quad \dots, \mathbf{X}(P-1); \alpha, \mathbf{A}, \mathbf{Q}) \\ &= f_i(\alpha, \mathbf{A}, \mathbf{Q}) \\ &= \left[ \frac{1}{\pi^J |\mathbf{Q}|} \exp \{ -\text{tr}(\mathbf{Q}^{-1} \mathbf{Q}_i(\mathbf{A})) \} \right]^{(K+1)(N-P)}, \end{aligned} \quad (\text{A.3})$$

where in the first equality we dropped the dependence on the observed signals for notational brevity,

$$\mathbf{Q}_i(\alpha, \mathbf{A}) = \frac{1}{(K+1)(N-P)} \sum_{n=P}^{N-1} \sum_{k=0}^K \varepsilon_k(n) \varepsilon_k^H(n), \quad (\text{A.4})$$

and we reiterate that  $\alpha = 0$  for  $i = 0$ ,  $\alpha \neq 0$  for  $i = 1$ , and  $\varepsilon_0(n)$  depends on  $\alpha$  as shown in (3.8) or (3.11).

The Rao test requires the ML estimates of the nuisance parameters under  $H_0$ . Henceforth, we only consider the case  $i = 0$ . Taking the derivative of  $\ln f_0(\mathbf{A}, \mathbf{Q})$  with respect to  $\mathbf{Q}$  and equating the result to zero produces the ML estimates of  $\mathbf{Q}(\mathbf{A})$  given  $\mathbf{A}$ :<sup>1</sup>

$$\hat{\mathbf{Q}}_{\text{ML}}(\mathbf{A}) = \mathbf{Q}_0(\mathbf{A}). \quad (\text{A.5})$$

Substituting  $\hat{\mathbf{Q}}_{\text{ML}}(\mathbf{A})$  into  $f_0(\mathbf{A}, \mathbf{Q})$ , we have

$$\max_{\mathbf{Q}} f_0(\mathbf{A}, \mathbf{Q}) = \left[ \frac{1}{(e\pi)^J |\mathbf{Q}_0(\mathbf{A})|} \right]^{(K+1)(N-P)}. \quad (\text{A.6})$$

Next, we determine the ML estimates of  $\mathbf{A}$ . Since maximizing  $f_0(\mathbf{A})$  is equivalent to minimizing  $|\hat{\mathbf{Q}}_{\text{ML}}(\mathbf{A})|$ , or  $|\mathbf{Q}_0(\mathbf{A})|$ , the ML estimate of  $\mathbf{A}$  can be obtained by minimizing  $|\mathbf{Q}_0(\mathbf{A})|$  with respect to  $\mathbf{A}$ . We next expand the matrix as follows:

$$\begin{aligned} & (K+1)(N-P) \mathbf{Q}_0(\mathbf{A}) \\ &= \hat{\mathbf{R}}_{xx} + \mathbf{A}^H \hat{\mathbf{R}}_{yx} + \hat{\mathbf{R}}_{yx}^H \mathbf{A} + \mathbf{A}^H \hat{\mathbf{R}}_{yy} \mathbf{A} \\ &= \left( \mathbf{A}^H + \hat{\mathbf{R}}_{yx}^H \hat{\mathbf{R}}_{yy}^{-1} \right) \hat{\mathbf{R}}_{yy} \left( \mathbf{A}^H + \hat{\mathbf{R}}_{yx}^H \hat{\mathbf{R}}_{yy}^{-1} \right)^H \\ & \quad + \hat{\mathbf{R}}_{xx} - \hat{\mathbf{R}}_{yx}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{R}}_{yx}, \end{aligned} \quad (\text{A.7})$$

where the correlation matrices are defined in (3.17)-(3.19). Since  $\hat{\mathbf{R}}_{yy}$  is non-negative definite and the remaining terms in (A.7) do not depend on  $\mathbf{A}$ , it follows that<sup>2</sup>

$$\mathbf{Q}_0(\mathbf{A}) \geq \mathbf{Q}_0(\mathbf{A})|_{\mathbf{A}=\hat{\mathbf{A}}}, \quad (\text{A.8})$$

<sup>1</sup>Since  $\alpha = 0$  for  $i = 0$ , the dependence on  $\alpha$  is dropped.

<sup>2</sup>For two non-negative definite matrices  $\mathbf{A}$  and  $\mathbf{B}$ , we have  $\mathbf{A} \geq \mathbf{B}$  if  $\mathbf{A} - \mathbf{B}$  is non-negative definite.

where

$$\hat{\mathbf{A}}^H = -\hat{\mathbf{R}}_{yx}^H \hat{\mathbf{R}}_{yy}^{-1}. \quad (\text{A.9})$$

When  $\mathbf{Q}_0(\mathbf{A})$  is minimized, the estimate  $\hat{\mathbf{A}}^H$  of  $\mathbf{A}^H$  will minimize any non-decreasing function including the determinant of  $\mathbf{Q}_0(\mathbf{A})$  [41]. Hence, the ML estimate  $\hat{\mathbf{A}}_{\text{ML}}^H$  of  $\mathbf{A}^H$  is given by (A.9) or (3.20), and  $\hat{\mathbf{Q}}_{\text{ML}}$  is given by (3.21), which is obtained by replacing  $\mathbf{A}^H$  in (A.7) with  $\hat{\mathbf{A}}_{\text{ML}}^H$ . The subscript “ML” is dropped in other parts of the paper for notational brevity.

## A.2 Derivation of the Parametric Rao Test

The composite hypothesis testing problem (3.3) involves a *signal parameter vector*  $\boldsymbol{\theta}_r = [\alpha_R, \alpha_I]^T = [\Re\{\alpha\}, \Im\{\alpha\}]^T$  and a *nuisance parameter vector*  $\boldsymbol{\theta}_s$  that includes all unknown parameters in  $\{\mathbf{A}^H(p)\}_{p=1}^P$  and  $\mathbf{Q}$ . The nuisance parameter vector  $\boldsymbol{\theta}_s$  may be written as  $\boldsymbol{\theta}_s = [\mathbf{q}_R^T, \mathbf{q}_I^T, \mathbf{a}_R^T, \mathbf{a}_I^T]^T$  with  $\mathbf{a}_R = \text{vec}(\Re\{\mathbf{A}^H\})$ ,  $\mathbf{a}_I = \text{vec}(\Im\{\mathbf{A}^H\})$ ,  $\mathbf{q}_R$  contains the diagonal elements in  $\mathbf{Q}$  and the real part of the elements below the diagonal, while  $\mathbf{q}_I$  contains the imaginary part of the elements below the diagonal (note that the spatial covariance matrix  $\mathbf{Q}$  is a Hermitian matrix). Let

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T, \boldsymbol{\theta}_s^T]^T. \quad (\text{A.10})$$

Observing that the nuisance parameters are the same under both hypotheses, we can write the parameter test as follows:

$$\begin{aligned} H_0 : \quad & \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s, \\ H_1 : \quad & \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_1}, \boldsymbol{\theta}_s, \end{aligned} \quad (\text{A.11})$$

where  $\boldsymbol{\theta}_{r_0} = [0, 0]^T$  and  $\boldsymbol{\theta}_{r_1} = \boldsymbol{\theta}_r = [\alpha_R, \alpha_I]^T$ . The PDF under  $H_0$  and the PDF under  $H_1$  differ only in the value of  $\boldsymbol{\theta}$ , and they are given by (see Appendix A.1):

$$f_i(\boldsymbol{\theta}) = \left[ \frac{1}{\pi^J |\mathbf{Q}|} \exp \left\{ -\text{tr}(\mathbf{Q}^{-1} \mathbf{Q}_i(\alpha, \mathbf{A})) \right\} \right]^{(K+1)(N-P)},$$

where  $\mathbf{Q}_i(\alpha, \mathbf{A})$  is defined in (A.4). The Rao test is given by [7]

$$\left. \frac{\partial \ln f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[ \mathbf{J}^{-1}(\tilde{\boldsymbol{\theta}}) \right]_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r} \left. \frac{\partial \ln f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{Rao}}, \quad (\text{A.12})$$

where  $\gamma_{\text{Rao}}$  denotes a corresponding threshold,

$$\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}_{r_0}^T, \hat{\boldsymbol{\theta}}_{s_0}^T]^T \quad (\text{A.13})$$

denotes the ML estimate of  $\boldsymbol{\theta}$  under  $H_0$ , and

$$\begin{aligned} & \left[ \mathbf{J}^{-1}(\boldsymbol{\theta}) \right]_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r} \\ & = \left[ \mathbf{J}_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r}(\boldsymbol{\theta}) - \mathbf{J}_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_s}(\boldsymbol{\theta}) \mathbf{J}_{\boldsymbol{\theta}_s, \boldsymbol{\theta}_s}^{-1}(\boldsymbol{\theta}) \mathbf{J}_{\boldsymbol{\theta}_s, \boldsymbol{\theta}_r}(\boldsymbol{\theta}) \right]^{-1}, \end{aligned} \quad (\text{A.14})$$

which is related to the *Fisher information matrix* (FIM), given by [7]

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r}(\boldsymbol{\theta}) & \mathbf{J}_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_s}(\boldsymbol{\theta}) \\ \mathbf{J}_{\boldsymbol{\theta}_s, \boldsymbol{\theta}_r}(\boldsymbol{\theta}) & \mathbf{J}_{\boldsymbol{\theta}_s, \boldsymbol{\theta}_s}(\boldsymbol{\theta}) \end{bmatrix}. \quad (\text{A.15})$$

Hence, the problem boils down to finding the ML estimates of the nuisance parameters under  $H_0$ , which have been obtained in Appendix A.1, and evaluating the first order derivatives of the log likelihood and the FIM at the ML estimates of the nuisance parameters. The latter task is worked out next.

The FIM is block diagonal. To see this, let  $q_{R_i}$ ,  $q_{I_i}$ ,  $a_{R_i}$ , and  $a_{I_i}$  denote the  $i$ -th element of  $\mathbf{q}_R$ ,  $\mathbf{q}_I$ ,  $\mathbf{a}_R$  and  $\mathbf{a}_I$ , respectively. The first partial derivative of the log likelihood  $\ln f$  with respect to (w.r.t.)  $\alpha_R$  is

$$\frac{\partial \ln f}{\partial \alpha_R} = \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{Q}^{-1} \boldsymbol{\varepsilon}_0(n) + \sum_{n=P}^{N-1} \boldsymbol{\varepsilon}_0^H(n) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n). \quad (\text{A.16})$$

The second partial derivative of  $\ln f$  w.r.t.  $\alpha_R$  and  $q_{R_i}$  becomes

$$\frac{\partial^2 \ln f}{\partial \alpha_R \partial q_{R_i}} = -2\Re \left\{ \sum_{n=P}^{N-1} \boldsymbol{\varepsilon}_0^H(n) \mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial q_{R_i}} \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n) \right\}. \quad (\text{A.17})$$

Likewise, we have

$$\frac{\partial^2 \ln f}{\partial \alpha_R \partial q_{I_i}} = -2\Im \left\{ \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial q_{I_i}} \mathbf{Q}^{-1} \boldsymbol{\varepsilon}_0(n) \right\} \quad (\text{A.18})$$

and

$$\begin{aligned} & \frac{\partial^2 \ln f}{\partial \alpha_R \partial a_{R_i}} \\ &= 2 \sum_{n=P}^{N-1} \sum_{p=1}^P \Re \left\{ \alpha \boldsymbol{\varepsilon}_0^H(n) \mathbf{Q}^{-1} \frac{\partial \mathbf{A}^H(p)}{\partial a_{R_i}} \mathbf{s}(n-p) \right. \\ & \quad \left. + \tilde{\mathbf{s}}^H(n) \mathbf{Q}^{-1} \frac{\partial \mathbf{A}^H(p)}{\partial a_{R_i}} [\mathbf{x}_0(n-p) - \alpha \mathbf{s}(n-p)] \right\}. \end{aligned} \quad (\text{A.19})$$

Since  $E[\mathbf{x}_0(n) - \alpha \mathbf{s}(n)] = E[\mathbf{x}_0(n-p) - \alpha \mathbf{s}(n-p)] = 0$  and  $E[\boldsymbol{\varepsilon}_0(n)] = 0$ , taking the expectation in (A.17)–(A.19) yields

$$E \left[ \frac{\partial^2 \ln f}{\partial \alpha_R \partial q_{R_i}} \right] = E \left[ \frac{\partial^2 \ln f}{\partial \alpha_R \partial q_{I_i}} \right] = E \left[ \frac{\partial^2 \ln f}{\partial \alpha_R \partial a_{R_i}} \right] = 0. \quad (\text{A.20})$$

In a similar way, we can show

$$\begin{aligned} & E \left[ \frac{\partial^2 \ln f}{\partial \alpha_I \partial q_{R_i}} \right] = E \left[ \frac{\partial^2 \ln f}{\partial \alpha_I \partial q_{I_i}} \right] = E \left[ \frac{\partial^2 \ln f}{\partial \alpha_R \partial a_{I_i}} \right] \\ &= E \left[ \frac{\partial^2 \ln f}{\partial \alpha_I \partial a_{R_i}} \right] = E \left[ \frac{\partial^2 \ln f}{\partial \alpha_I \partial a_{I_i}} \right] = 0. \end{aligned} \quad (\text{A.21})$$

Summarizing the above calculations, we have

$$\mathbf{J}_{\theta_r, \theta_s}(\boldsymbol{\theta}) = \mathbf{0}, \quad \mathbf{J}_{\theta_s, \theta_r}(\boldsymbol{\theta}) = \mathbf{0}, \quad (\text{A.22})$$

which implies that the FIM is block diagonal. It follows that

$$\left[ \mathbf{J}^{-1}(\tilde{\boldsymbol{\theta}}) \right]_{\theta_r, \theta_r} = \mathbf{J}_{\theta_r, \theta_r}^{-1}(\tilde{\boldsymbol{\theta}}). \quad (\text{A.23})$$

Hence, we only need to compute  $\mathbf{J}_{\theta_r, \theta_r}^{-1}(\tilde{\boldsymbol{\theta}})$ , which is obtained next. The second partial derivative of  $\ln f$  w.r.t.  $\alpha_R$  is

$$\frac{\partial^2 \ln f}{\partial \alpha_R^2} = -2 \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n). \quad (\text{A.24})$$

Likewise, we have

$$\frac{\partial^2 \ln f}{\partial \alpha_I^2} = -2 \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n) \quad (\text{A.25})$$

and

$$\frac{\partial^2 \ln f}{\partial \alpha_R \partial \alpha_I} = \frac{\partial^2 \ln f}{\partial \alpha_I \partial \alpha_R} = 0. \quad (\text{A.26})$$

As a result, we have the FIM associated with the signal parameter vector:

$$\mathbf{J}_{\theta_r, \theta_r}(\boldsymbol{\theta}) = 2 \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (\text{A.27})$$

Finally, by inverting the matrix (A.27) and replacing  $\boldsymbol{\theta}$  with  $\tilde{\boldsymbol{\theta}}$  which is the ML estimate of  $\boldsymbol{\theta}$  under  $H_0$ , we have

$$\mathbf{J}_{\theta_r, \theta_r}^{-1}(\tilde{\boldsymbol{\theta}}) = \frac{1}{2 \sum_{n=P}^{N-1} \hat{\mathbf{s}}^H(n) \hat{\mathbf{Q}}^{-1} \hat{\mathbf{s}}(n)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (\text{A.28})$$

where  $\hat{\mathbf{Q}}$  is the ML estimate of the spatial covariance matrix in (3.21), and  $\hat{\mathbf{s}}(n)$  is the temporally whitened steering vector in (3.13). Moreover, since  $\boldsymbol{\varepsilon}_0(n)|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \hat{\mathbf{x}}_0(n)$ , we have

$$\left. \frac{\partial \ln f}{\partial \alpha_R} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \sum_{n=P}^{N-1} \left\{ \hat{\mathbf{x}}_0^H(n) \hat{\mathbf{Q}}^{-1} \hat{\mathbf{s}}(n) + \hat{\mathbf{s}}^H(n) \hat{\mathbf{Q}}^{-1} \hat{\mathbf{x}}_0(n) \right\} \quad (\text{A.29})$$

$$\left. \frac{\partial \ln f}{\partial \alpha_I} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = j \sum_{n=P}^{N-1} \left\{ \hat{\mathbf{x}}_0^H(n) \hat{\mathbf{Q}}^{-1} \hat{\mathbf{s}}(n) + \hat{\mathbf{s}}^H(n) \hat{\mathbf{Q}}^{-1} \hat{\mathbf{x}}_0(n) \right\}. \quad (\text{A.30})$$

Using (A.28)-(A.30) in (A.12) yields the parametric Rao test, which is shown in (7.14).

### A.3 Asymptotic Distribution of the Parametric Rao Test Statistic

The Rao test is known to have the same asymptotic performance as the GLRT. Using the asymptotic results for the GLRT [7], the asymptotic distribution of our parametric Rao test statistic is given by

$$T_{\text{Rao}} \stackrel{a}{\sim} \begin{cases} \chi_2^2, & \text{under } H_0, \\ \chi_2'^2(\lambda), & \text{under } H_1, \end{cases} \quad (\text{A.31})$$

where  $\chi_2^2$  denotes the central Chi-squared distribution with 2 degrees of freedom and  $\chi_2'^2(\lambda)$  the non-central Chi-squared distribution with 2 degrees of freedom and non-centrality parameter  $\lambda$ :

$$\lambda = (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left( [\mathbf{J}^{-1}([\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s])]_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r} \right)^{-1} (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0}). \quad (\text{A.32})$$

Using the observations  $\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0} = [\alpha_R, \alpha_I]$  and (cf. (A.27))

$$[\mathbf{J}^{-1}([\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s])]_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r} = \frac{1}{2 \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n)} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (\text{A.33})$$

we have the asymptotic distribution of the parametric Rao test statistic as shown in (3.22).

### A.4 Performance of the MF and AMF Detectors

The performance of the MF and AMF detectors can be computed analytically. In this appendix, we include a brief summary of their performance for easy reference.

Consider the MF detector (8.3) first. Let  $\mathbf{R}^{-1/2}$  be the square-root of the space-time covariance matrix  $\mathbf{R}$ . Define  $\tilde{\mathbf{s}} = \mathbf{R}^{-1/2} \mathbf{s}$  and  $\tilde{\mathbf{x}}_0 = \mathbf{R}^{-1/2} \mathbf{x}_0$ , which are the spatially and temporally whitened steering vector and test signal, respectively. Since the rank of  $\tilde{\mathbf{s}} \tilde{\mathbf{s}}^H$  is one, we have the following eigen decomposition:

$$\tilde{\mathbf{s}} \tilde{\mathbf{s}}^H = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^H, \quad (\text{A.34})$$

where  $\boldsymbol{\Lambda} = \text{diag}(\tilde{\mathbf{s}}^H \tilde{\mathbf{s}}, 0, \dots, 0)$  and  $\mathbf{U}^H \mathbf{U} = \mathbf{I}$ . Let  $\bar{\mathbf{x}}_0 = \mathbf{U}^H \tilde{\mathbf{x}}_0$  and  $\bar{\mathbf{s}} = \mathbf{U}^H \tilde{\mathbf{s}}$ , which are rotated versions of  $\tilde{\mathbf{x}}_0$  and  $\tilde{\mathbf{s}}$ , respectively. Then, the test statistic can be written as

$$T_{\text{MF}} = \frac{|\tilde{\mathbf{s}}^H \tilde{\mathbf{x}}_0|^2}{\tilde{\mathbf{s}}^H \tilde{\mathbf{s}}} = \frac{\tilde{\mathbf{x}}_0^H \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^H \tilde{\mathbf{x}}_0}{\tilde{\mathbf{s}}^H \tilde{\mathbf{s}}} = \bar{x}_{0,1}^* \bar{x}_{0,1} = |\bar{x}_{0,1}|^2, \quad (\text{A.35})$$

where  $\bar{x}_{0,1}$  and  $\bar{s}_1$  denotes the first element of  $\bar{\mathbf{x}}_0$  and  $\bar{\mathbf{s}}$ , respectively. It is clear from Assumptions **AS1** to **AS3** in Section 7.2 that  $\bar{x}_{0,1}$  is a complex Gaussian variable:  $\bar{x}_{0,1} \sim \mathcal{CN}(\alpha \bar{s}_1, 1)$  with  $\alpha = 0$  under  $H_0$  and  $\alpha \neq 0$  under  $H_1$ . Hence,  $2T_{\text{MF}} = 2|\bar{x}_{0,1}|^2$  has a central Chi-squared distribution with 2 degrees of freedom under  $H_0$  and, respectively, a non-central Chi-squared distribution with 2 degrees of freedom and a non-centrality parameter  $\lambda_{\text{MF}} = 2|\alpha \bar{s}_1|^2$  under  $H_1$ .



It is noted that the distribution of the MF test statistic is similar to that of the parametric Rao test statistic with the only difference of the non-centrality parameter under  $H_1$ . Hence, the false alarm and detection probabilities can be similarly computed as in (3.27) and (3.28).

The performance of the AMF detector (8.5) was analyzed in [12], which is summarized below. The density of a loss factor  $\rho$ , which was defined in (25) of [12], is given by

$$f(\rho) = f_\beta(\rho; L - 1, JN - 1), \quad (\text{A.36})$$

where  $L = K - JN + 1$  and the central Beta density function is

$$f_\beta(x; n, m) = \frac{(n + m - 1)!}{(n - 1)!(m - 1)!} x^{n-1} (1 - x)^{m-1}. \quad (\text{A.37})$$

The probability of false alarm is given by

$$P_{\text{f,AMF}} = \int_0^1 \frac{f_\beta(\rho; L - 1, JN - 1)}{(1 + \eta\rho)^L} d\rho, \quad (\text{A.38})$$

where  $\eta = \gamma_{\text{Kelly}} / (1 - \gamma_{\text{Kelly}})$  and  $\gamma_{\text{Kelly}}$  is the test threshold of Kelly's GLRT (8.6). Meanwhile, the probability of detection is given by

$$\begin{aligned} P_{\text{d,AMF}} = 1 - \int_0^1 \frac{1}{(1 + \eta\rho)^L} \sum_{m=1}^L \binom{L}{m} \\ \times (\eta\rho)^m G_m\left(\frac{\xi\rho}{1 + \eta\rho}\right) f(\rho) d\rho, \end{aligned} \quad (\text{A.39})$$

where  $\xi = \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}$  and  $G_m(\cdot)$  is the incomplete Gamma function given by

$$G_m(y) = e^{-y} \sum_{k=0}^{m-1} \frac{y^k}{k!}. \quad (\text{A.40})$$

The integrals can be computed by numerical integration.

# Appendix B

## Parametric GLRT

### B.1 ML Parameter Estimation

In this appendix, we develop the ML parameter estimators under both hypotheses. Recall that the likelihood functions under both hypotheses differ only in the value of  $\alpha$ , that is,  $\alpha = 0$  under  $H_0$  and  $\alpha \neq 0$  under  $H_1$ . We will show that the ML estimates under  $H_0$  can be obtained by setting  $\alpha = 0$  in the ML estimates under  $H_1$ .

Let  $\tilde{\mathbf{x}}_k(n; \mathbf{A})$  denote the temporally whitened version of  $\mathbf{x}_k(n)$ :

$$\tilde{\mathbf{x}}_k(n; \mathbf{A}) = \mathbf{x}_k(n) + \sum_{p=1}^P \mathbf{A}^H(p) \mathbf{x}_k(n-p). \quad (\text{B.1})$$

Conditioned on the first  $P$  values  $\{\mathbf{x}_k(n)\}_{n=0}^{P-1}$ ,  $k = 0, 1, \dots, K$ , the log-likelihood function is proportional to (within an additive constant) [37]

$$\begin{aligned} -L \ln |\mathbf{Q}| - \sum_{k=1}^K \sum_{n=P}^{N-1} \tilde{\mathbf{x}}_k^H(n; \mathbf{A}) \mathbf{Q}^{-1} \tilde{\mathbf{x}}_k(n; \mathbf{A}) \\ - \sum_{n=P}^{N-1} \{\tilde{\mathbf{x}}_0(n; \mathbf{A}) - \alpha \tilde{\mathbf{s}}(n; \mathbf{A})\}^H \mathbf{Q}^{-1} \\ \times \{\tilde{\mathbf{x}}_0(n; \mathbf{A}) - \alpha \tilde{\mathbf{s}}(n; \mathbf{A})\}. \end{aligned} \quad (\text{B.2})$$

It is noted that for the large-sample case, the likelihood function can be well approximated by the above conditional distribution [30]. We therefore use (B.2) for ML estimation. Taking the derivative of (B.2) with respect to  $\mathbf{Q}$  and equating it to zero produce the ML estimates of  $\mathbf{Q}$  conditioned on  $\alpha$  and  $\mathbf{A}$ :

$$\begin{aligned} \hat{\mathbf{Q}}(\alpha, \mathbf{A}) = \frac{1}{L} \sum_{n=P}^{N-1} \sum_{k=1}^K \left[ \tilde{\mathbf{x}}_k(n; \mathbf{A}) \tilde{\mathbf{x}}_k^H(n; \mathbf{A}) + \right. \\ \left. \{\tilde{\mathbf{x}}_0(n; \mathbf{A}) - \alpha \tilde{\mathbf{s}}(n; \mathbf{A})\} \{\tilde{\mathbf{x}}_0(n; \mathbf{A}) - \alpha \tilde{\mathbf{s}}(n; \mathbf{A})\}^H \right]. \end{aligned} \quad (\text{B.3})$$

Substituting the above  $\hat{\mathbf{Q}}(\alpha, \mathbf{A})$  back in (B.2), we find that maximizing (B.2) reduces to minimizing  $|\hat{\mathbf{Q}}(\alpha, \mathbf{A})|$ . Therefore, the ML estimates of  $\alpha$  and  $\mathbf{A}$  can be obtained by minimizing  $|\hat{\mathbf{Q}}(\alpha, \mathbf{A})|$  with respect to  $\alpha$  and  $\mathbf{A}$ . In turn, we can get the ML estimate of  $\mathbf{Q}$  by replacing  $\alpha$  and  $\mathbf{A}$  with their ML estimates in (B.3). Next, observe that

$$\begin{aligned}
L\hat{\mathbf{Q}}(\alpha, \mathbf{A}) &= \hat{\mathbf{R}}_{xx}(\alpha) + \mathbf{A}^H \hat{\mathbf{R}}_{yx}(\alpha) + \hat{\mathbf{R}}_{yx}^H(\alpha) \mathbf{A} + \mathbf{A}^H \hat{\mathbf{R}}_{yy}(\alpha) \mathbf{A} \\
&= \left( \mathbf{A}^H + \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \right) \hat{\mathbf{R}}_{yy}(\alpha) \\
&\quad \times \left( \mathbf{A}^H + \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \right)^H \\
&\quad + \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha) \hat{\mathbf{R}}_{yx}(\alpha),
\end{aligned} \tag{B.4}$$

where the  $\alpha$ -dependent correlation matrices are defined in (4.10)-(4.12). Since  $\hat{\mathbf{R}}_{yy}(\alpha)$  is non-negative definite and the remaining terms in (B.4) do not depend on  $\mathbf{A}$ , it follows that<sup>1</sup>

$$\hat{\mathbf{Q}}(\alpha, \mathbf{A}) \geq \hat{\mathbf{Q}}(\alpha, \mathbf{A})|_{\mathbf{A}=\hat{\mathbf{A}}(\alpha)}, \tag{B.5}$$

where

$$\hat{\mathbf{A}}^H(\alpha) = -\hat{\mathbf{R}}_{yx}^H(\alpha) \hat{\mathbf{R}}_{yy}^{-1}(\alpha). \tag{B.6}$$

When  $\hat{\mathbf{Q}}(\alpha, \mathbf{A})$  is minimized, the estimate  $\hat{\mathbf{A}}(\alpha)$  of  $\mathbf{A}$  will minimize any non-decreasing function including the determinant of  $\hat{\mathbf{Q}}(\alpha, \mathbf{A})$  [41]. It should be noted that in finding the estimate of  $\mathbf{A}$ , we did not impose the constraint that the underlying AR process is stable for the sake of obtaining a simple solution. Hence, the unconstrained ML estimate of  $\mathbf{A}$  and  $\mathbf{Q}$  conditioned on  $\alpha$  are given by (4.17) and (4.18), respectively.

Replacing  $\mathbf{A}$  in (B.4) by  $\hat{\mathbf{A}}(\alpha)$  followed by minimizing  $|\hat{\mathbf{Q}}(\alpha, \hat{\mathbf{A}}(\alpha))|$  yields the ML amplitude estimator of  $\alpha$  given by (4.9). Once the ML estimate  $\hat{\alpha}_{\text{ML}}$  of the signal amplitude  $\alpha$  is obtained, substituting  $\hat{\alpha}_{\text{ML}}$  in (4.17) and (4.18) yields the ML estimates of  $\mathbf{A}$  and  $\mathbf{Q}$  under  $H_1$ , which are given by (4.15) and (4.16), respectively.

Since  $\alpha = 0$  under  $H_0$ , substituting  $\alpha = 0$  in (B.5) and (B.6) leads to the ML estimates of  $\mathbf{A}$  and  $\mathbf{Q}$  under  $H_0$ , which are given by (4.20) and (4.21), respectively.

---

<sup>1</sup>For two non-negative definite matrices  $\mathbf{A}$  and  $\mathbf{B}$ , we have  $\mathbf{A} \geq \mathbf{B}$  if  $\mathbf{A} - \mathbf{B}$  is non-negative definite [41].

## B.2 Derivation of the AML Estimator

Using definitions in (4.26)-(4.29), (4.22) can be written as

$$\begin{aligned}
& \left| \left( \hat{\mathbf{X}}_0 - \alpha \hat{\mathbf{S}} \right) \left( \hat{\mathbf{X}}_0 - \alpha \hat{\mathbf{S}} \right)^H + \sum_{k=1}^K \hat{\mathbf{X}}_k \hat{\mathbf{X}}_k^H \right| \\
&= \left| \left( \hat{\mathbf{X}}_0 - \alpha \hat{\mathbf{S}} \right) (\mathbf{P} + \mathbf{P}^\perp) \left( \hat{\mathbf{X}}_0 - \alpha \hat{\mathbf{S}} \right)^H + \sum_{k=1}^K \hat{\mathbf{X}}_k \hat{\mathbf{X}}_k^H \right| \\
&= \left| \left( \hat{\mathbf{X}}_0 \mathbf{P} - \alpha \hat{\mathbf{S}} \right) \left( \hat{\mathbf{X}}_0 \mathbf{P} - \alpha \hat{\mathbf{S}} \right)^H + \boldsymbol{\Psi} \right| \\
&= \left| \left( \hat{\mathbf{X}}_0 \mathbf{P} - \alpha \hat{\mathbf{S}} \right) \left( \hat{\mathbf{X}}_0 \mathbf{P} - \alpha \hat{\mathbf{S}} \right)^H \boldsymbol{\Psi}^{-1} + \mathbf{I} \right| |\boldsymbol{\Psi}|.
\end{aligned} \tag{B.7}$$

Next, observe that minimizing

$$C_2(\alpha) = \left| \left( \hat{\mathbf{X}}_0 \mathbf{P} - \alpha \hat{\mathbf{S}} \right) \left( \hat{\mathbf{X}}_0 \mathbf{P} - \alpha \hat{\mathbf{S}} \right)^H \boldsymbol{\Psi}^{-1} + \mathbf{I} \right|, \tag{B.8}$$

is asymptotically equivalent to minimizing [39, 41]:

$$C_3(\alpha) = \text{tr} \left\{ \left( \hat{\mathbf{X}}_0 \mathbf{P} - \alpha \hat{\mathbf{S}} \right)^H \boldsymbol{\Psi}^{-1} \left( \hat{\mathbf{X}}_0 \mathbf{P} - \alpha \hat{\mathbf{S}} \right) \right\}. \tag{B.9}$$

which is a quadratic function in  $\alpha$ . Minimizing (B.9) with respect to  $\alpha$  leads to the AML estimate  $\hat{\alpha}_{\text{AML}}$  given by (4.25) (also see [39]).

## B.3 Unbiasedness and Consistency of the LS Estimator

First, note that

$$E[\hat{\alpha}_{\text{LS}}] = E \left[ \frac{\mathbf{s}^H \mathbf{x}_0}{\mathbf{s}^H \mathbf{s}} \right] = E \left[ \alpha + \frac{\mathbf{s}^H \mathbf{d}_0}{\mathbf{s}^H \mathbf{s}} \right] = \alpha, \tag{B.10}$$

which indicates that the LS estimator is unbiased. Moreover, the variance is given by

$$\text{var}[\hat{\alpha}_{\text{LS}}] = E \left[ \left( \frac{\mathbf{s}^H \mathbf{d}_0}{\mathbf{s}^H \mathbf{s}} \right) \left( \frac{\mathbf{s}^H \mathbf{d}_0}{\mathbf{s}^H \mathbf{s}} \right)^H \right] = \frac{\mathbf{s}^H \mathbf{R} \mathbf{s}}{(\mathbf{s}^H \mathbf{s})^2}. \tag{B.11}$$

Next, we show that the variance vanishes as the number of observations  $N$  increases. Note that the numerator can be written as

$$\mathbf{s}^H \mathbf{R} \mathbf{s} = \mathbf{s}^H \mathbf{R}^{1/2} \mathbf{R}^{1/2} \mathbf{s}, \tag{B.12}$$

where  $\mathbf{R}^{1/2}$  denotes the Hermitian square-root of  $\mathbf{R}$ . It is known that multiplying by  $\mathbf{R}^{1/2}$  is a *coloring* linear transform. Under Assumption **AS4**, such a coloring transform using the square-root of the joint space-time covariance matrix  $\mathbf{R}$  is asymptotically equivalent to a cascade of an

AR filter, which performs temporal coloring, followed by a spatial coloring filter [30]. As such, (B.12) can be approximated (for large  $N$ ) as

$$\mathbf{s}^H \mathbf{R} \mathbf{s} \approx \sum_{n=0}^{N-1} \check{\mathbf{s}}^H(n) \mathbf{Q} \check{\mathbf{s}}(n), \quad (\text{B.13})$$

where  $\check{\mathbf{s}}(n) \in \mathbb{C}^{J \times 1}$  denotes the output of the multichannel AR filter as specified in **AS4**, given the input signal  $\{\mathbf{s}(n)\}_{n=0}^{N-1}$ .

Let the eigenvalue decomposition of  $\mathbf{Q}$  be expressed as:  $\mathbf{Q} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$ , where  $\mathbf{\Lambda}$  is a diagonal matrix containing all eigenvalues and  $\mathbf{U}$  is composed of the corresponding eigenvectors. Let  $\lambda_{\max}$  denote the largest eigenvalue of  $\mathbf{Q}$ . We have

$$\begin{aligned} \sum_{n=0}^{N-1} \check{\mathbf{s}}^H(n) \mathbf{Q} \check{\mathbf{s}}(n) &\leq \lambda_{\max} \sum_{n=0}^{N-1} \check{\mathbf{s}}^H(n) \mathbf{U} \mathbf{U}^H \check{\mathbf{s}}(n) \\ &= \lambda_{\max} \sum_{n=0}^{N-1} \|\check{\mathbf{s}}(n)\|^2. \end{aligned} \quad (\text{B.14})$$

It follows that

$$\text{var}[\hat{\alpha}_{\text{LS}}] \leq \frac{\lambda_{\max} \sum_{n=0}^{N-1} \|\check{\mathbf{s}}(n)\|^2}{\left( \sum_{n=0}^{N-1} \|\mathbf{s}(n)\|^2 \right)^2}. \quad (\text{B.15})$$

since  $(\mathbf{s}^H \mathbf{s})^2 = \left( \sum_{n=0}^{N-1} \|\mathbf{s}(n)\|^2 \right)^2$ . Assuming that the AR filter is stable, we have (e.g., [81]):

$$\sum_{n=0}^{N-1} \|\check{\mathbf{s}}(n)\|^2 \leq C \sum_{n=0}^{N-1} \|\mathbf{s}(n)\|^2 \quad (\text{B.16})$$

for some bounded constant  $C$ . Hence, for a given AR filter and spatial covariance  $\mathbf{Q}$ , the right-hand side of (B.15) vanishes as  $N$  goes to infinity. This proves that the LS amplitude estimate is statistically consistent.

## B.4 Derivation of CRB

Let  $\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T, \boldsymbol{\theta}_s^T]^T$ , where  $\boldsymbol{\theta}_r = [\Re\{\alpha\}, \Im\{\alpha\}]^T$ , and  $\boldsymbol{\theta}_s$  contains all nuisance parameters in  $\{\mathbf{A}^H(p)\}_{p=1}^P$  and  $\mathbf{Q}$ . It is shown in [37] that the Fisher information matrix (FIM) for  $\boldsymbol{\theta}$  is block diagonal with respect to  $\boldsymbol{\theta}_r$  and  $\boldsymbol{\theta}_s$ . Therefore, the CRB for the signal amplitude estimate is given by the FIM associated with  $\boldsymbol{\theta}_r$ , which is given by [37]

$$\mathbf{J}(\boldsymbol{\theta})_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r} = \left[ 2 \sum_{n=p}^{N-1} \tilde{\mathbf{s}}^H(n; \mathbf{A}) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n; \mathbf{A}) \right] \mathbf{I}_2. \quad (\text{B.17})$$

By inverting (B.17) and using  $\text{CRB}(\alpha) = \text{CRB}(\Re\{\alpha\}) + \text{CRB}(\Im\{\alpha\})$ , we have the CRB given by (4.32).

## B.5 Asymptotic Distribution of the Parametric GLRT Statistic

Using the asymptotic results for the GLRT [7], the asymptotic distribution of our parametric GLRT statistic is given by

$$T_{\text{GLRT}} \stackrel{a}{\sim} \begin{cases} \chi_2^2, & \text{under } H_0, \\ \chi_2'^2(\lambda), & \text{under } H_1, \end{cases} \quad (\text{B.18})$$

where the non-centrality parameter  $\lambda$  is given by

$$\lambda = (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0})^T \left( [\mathbf{J}^{-1}([\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s])]_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r} \right)^{-1} (\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0}), \quad (\text{B.19})$$

where  $\boldsymbol{\theta}_{r_0}$  and  $\boldsymbol{\theta}_{r_1}$  are  $\boldsymbol{\theta}_r$  under  $H_0$  and  $H_1$ , respectively;  $[\mathbf{J}^{-1}([\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s])]_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r}$  is the  $2 \times 2$  upper-left partition of  $\mathbf{J}^{-1}([\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s])$ . Using the observations  $\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0} = [\alpha_R, \alpha_I]$  and (cf. (B.17))

$$\begin{aligned} & [\mathbf{J}^{-1}([\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s])]_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r} \\ &= \left[ 2 \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n; \mathbf{A}) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n; \mathbf{A}) \right]^{-1} \mathbf{I}_2, \end{aligned} \quad (\text{B.20})$$

we have the asymptotic distribution of the parametric GLRT statistic as shown in (4.36).

# Appendix C

## A Simplified Parametric GLRT

### C.1 Derivation of (5.23)

Starting from (5.14), the determinant of  $\hat{\mathbf{R}}(\alpha)$  can be written as

$$\begin{aligned}
 & \left| (\mathbf{X}_0 - \alpha \mathbf{S}) (\mathbf{X}_0 - \alpha \mathbf{S})^H + \sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^H \right| \\
 &= \left| (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S}) (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S})^H \right. \\
 & \quad \left. + \mathbf{X}_0 \mathbf{P}_S^\perp \mathbf{X}_0^H + \sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^H \right| \\
 &= \left| (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S}) (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S})^H \hat{\mathbf{R}}_X^{-1} + \mathbf{I} \right| \cdot \left| \hat{\mathbf{R}}_X \right|. \tag{C.1}
 \end{aligned}$$

Consider the idempotent matrices  $\mathbf{P}_S$  and  $\mathbf{P}_S^\perp$  and assuming the number of sample data is large enough, i.e.,  $N \gg 1$ , we have

$$\text{rank}(\mathbf{P}_S) \leq J(P+1), \text{ and } \text{rank}(\mathbf{P}_S^\perp) \geq N - P \tag{C.2}$$

where  $\text{rank}(\cdot)$  denotes the rank of a matrix. Then, we have [39]

$$\begin{aligned}
 & (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S}) (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S})^H \hat{\mathbf{R}}_X^{-1} \\
 &= O\left(\frac{1}{N - P}\right) \ll 1. \tag{C.3}
 \end{aligned}$$

Let  $\{\lambda_m\}_{m=1}^M$  denote the eigenvalues of the matrix (C.3), which satisfies  $0 \leq \lambda_m \ll 1$  according to (C.3). Then

$$\begin{aligned}
& \left| (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S}) (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S})^H \hat{\mathbf{R}}_X^{-1} + \mathbf{I} \right| \\
&= \prod_{m=1}^M (1 + \lambda_m) \\
&\stackrel{a}{\approx} 1 + \sum_{m=1}^M \lambda_m \\
&= 1 + \text{tr} \left[ (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S})^H \hat{\mathbf{R}}_X^{-1} (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S}) \right], \tag{C.4}
\end{aligned}$$

where the approximation (a) holds in a first-order sense. Similarly, the determinant of  $\hat{\mathbf{R}}_{yy}(\alpha)$  can be expressed as

$$\begin{aligned}
& \left| \hat{\mathbf{R}}_{yy}(\alpha) \right| \\
&= \left| (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T}) (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T})^H \hat{\mathbf{R}}_Y^{-1} + \mathbf{I} \right| \cdot \left| \hat{\mathbf{R}}_Y \right|, \tag{C.5}
\end{aligned}$$

and

$$\begin{aligned}
& \left| (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T}) (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T})^H \hat{\mathbf{R}}_Y^{-1} + \mathbf{I} \right| \\
&\approx 1 + \text{tr} \left\{ (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T})^H \hat{\mathbf{R}}_Y^{-1} (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T}) \right\}. \tag{C.6}
\end{aligned}$$

Then, combining (C.1), (C.4)-(C.6) and ignoring the items independent of  $\alpha$  result in the asymptotically equivalent expression in (5.23).

## C.2 Derivation of the Amplitude Estimator

Following Appendix I and noting that

$$\text{tr} \left\{ (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S})^H \hat{\mathbf{R}}_X^{-1} (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S}) \right\} \ll 1, \tag{C.7}$$

and

$$\text{tr} \left\{ (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T})^H \hat{\mathbf{R}}_Y^{-1} (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T}) \right\} \ll 1, \tag{C.8}$$

we can approximate (5.23) as

$$\begin{aligned}
& F_1(\alpha) \\
&= \text{tr} \left\{ (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S})^H \hat{\mathbf{R}}_X^{-1} (\mathbf{X}_0 \mathbf{P}_S - \alpha \mathbf{S}) \right\} \\
&\quad - \text{tr} \left\{ (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T})^H \hat{\mathbf{R}}_Y^{-1} (\mathbf{Y}_0 \mathbf{P}_T - \alpha \mathbf{T}) \right\}, \tag{C.9}
\end{aligned}$$



where the approximation  $\ln(1+x) \approx x$ , for  $x \ll 1$ , was invoked. The cost function  $F_1(\alpha)$  is a quadratic function with respect to  $\alpha$ . It is easy to show that minimizing (C.9) with respect to  $\alpha$  leads to the AML2 amplitude estimate  $\hat{\alpha}$  given by (5.27).

### C.3 Derivation of The New Parametric GLRT

Using the Schur complements, we can write (7.6) as

$$\begin{aligned}
& \ln \frac{|\hat{\mathbf{Q}}_{\text{ML},0}|}{|\hat{\mathbf{Q}}_{\text{ML},1}|} \\
&= \ln \frac{\left| \mathbf{X}_0 \mathbf{X}_0^H + \sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^H \right|}{\left| \mathbf{Y}_0 \mathbf{Y}_0^H + \sum_{k=1}^K \mathbf{Y}_k \mathbf{Y}_k^H \right|}} \\
&= \ln \frac{\left| (\mathbf{X}_0 - \hat{\alpha}_{\text{ML}} \mathbf{S}) (\mathbf{X}_0 - \hat{\alpha}_{\text{ML}} \mathbf{S})^H + \sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^H \right|}{\left| (\mathbf{Y}_0 - \hat{\alpha}_{\text{ML}} \mathbf{T}) (\mathbf{Y}_0 - \hat{\alpha}_{\text{ML}} \mathbf{T})^H + \sum_{k=1}^K \mathbf{Y}_k \mathbf{Y}_k^H \right|}} \\
&\propto \ln \frac{\left| \mathbf{X}_0 \mathbf{P}_S \mathbf{X}_0^H \hat{\mathbf{R}}_X^{-1} + \mathbf{I} \right|}{\left| \mathbf{Y}_0 \mathbf{P}_T \mathbf{Y}_0^H \hat{\mathbf{R}}_Y^{-1} + \mathbf{I} \right|}} \\
&= \ln \frac{\left| (\mathbf{X}_0 \mathbf{P}_S - \hat{\alpha}_{\text{ML}} \mathbf{S}) (\mathbf{X}_0 \mathbf{P}_S - \hat{\alpha}_{\text{ML}} \mathbf{S})^H \hat{\mathbf{R}}_X^{-1} + \mathbf{I} \right|}{\left| (\mathbf{Y}_0 \mathbf{P}_T - \hat{\alpha}_{\text{ML}} \mathbf{T}) (\mathbf{Y}_0 \mathbf{P}_T - \hat{\alpha}_{\text{ML}} \mathbf{T})^H \hat{\mathbf{R}}_Y^{-1} + \mathbf{I} \right|}}.
\end{aligned}$$

The RHS of the above equation can be further simplified using asymptotic approximations [see (C.4) and (C.5)]

$$\begin{aligned}
& \ln \frac{|\hat{\mathbf{Q}}_{\text{ML},0}|}{|\hat{\mathbf{Q}}_{\text{ML},1}|} \\
&\propto \text{tr} \left\{ \mathbf{P}_S^H \mathbf{X}_0^H \hat{\mathbf{R}}_X^{-1} \mathbf{X}_0 \mathbf{P}_S \right\} - \text{tr} \left\{ \mathbf{P}_T^H \mathbf{Y}_0^H \hat{\mathbf{R}}_Y^{-1} \mathbf{Y}_0 \mathbf{P}_T \right\} \\
&\quad - \text{tr} \left\{ (\mathbf{X}_0 \mathbf{P}_S - \hat{\alpha}_{\text{ML}} \mathbf{S})^H \hat{\mathbf{R}}_X^{-1} (\mathbf{X}_0 \mathbf{P}_S - \hat{\alpha}_{\text{ML}} \mathbf{S}) \right\} \\
&\quad + \text{tr} \left\{ (\mathbf{Y}_0 \mathbf{P}_T - \hat{\alpha}_{\text{ML}} \mathbf{T})^H \hat{\mathbf{R}}_Y^{-1} (\mathbf{Y}_0 \mathbf{P}_T - \hat{\alpha}_{\text{ML}} \mathbf{T}) \right\}.
\end{aligned}$$

Replacing the exact ML estimate with the AML2 amplitude estimation results in the approximate parametric GLRT.

$$\text{GLR} \approx \frac{\left| \text{tr} \left( \mathbf{S}^H \hat{\mathbf{R}}_X^{-1} \mathbf{X}_0 \right) - \text{tr} \left( \mathbf{T}^H \hat{\mathbf{R}}_Y^{-1} \mathbf{Y}_0 \right) \right|^2}{\text{tr} \left( \mathbf{S}^H \hat{\mathbf{R}}_X^{-1} \mathbf{S} \right) - \text{tr} \left( \mathbf{T}^H \hat{\mathbf{R}}_Y^{-1} \mathbf{T} \right)}, \quad (\text{C.10})$$

which is the matrix form of (5.28).

## C.4 Alternative Form of The Parametric GLRT

Let

$$\mathbf{S}_1 = [\mathbf{s}(P), \mathbf{s}(P+1), \dots, \mathbf{s}(N-1)] \in \mathbb{C}^{J \times (N-P)} \quad (\text{C.11})$$

and  $\mathbf{X}_{k,1} \in \mathbb{C}^{J \times (N-P)}$  is similarly defined. The matrix  $\mathbf{S}$  can be rewritten as  $\mathbf{S}^H = [\mathbf{T}^H, \mathbf{S}_1^H]$ , where  $\mathbf{T} \in \mathbb{C}^{JP \times (N-P)}$  is given by (5.21). By invoking the formula of the block matrix pseudo-inverse [82], we have

$$\begin{aligned} \mathbf{P}_S^\perp &= \mathbf{I} - [\mathbf{T}^H \mathbf{S}_1^H] \\ &\quad \times \left[ (\mathbf{T}^H)^\dagger - (\mathbf{T}^H)^\dagger \mathbf{S}_1^H (\mathbf{C}^\dagger + \mathbf{D}) \right], \end{aligned} \quad (\text{C.12})$$

where

$$\mathbf{C} = (\mathbf{I}_{N-P} - \mathbf{T}^H (\mathbf{T}^H)^\dagger) \mathbf{S}_1^H \in \mathbb{C}^{(N-P) \times J}, \quad (\text{C.13})$$

and

$$\begin{aligned} \mathbf{D} &= (\mathbf{I}_J - \mathbf{C}^\dagger \mathbf{C}) \left[ \mathbf{I}_J + (\mathbf{I}_J - \mathbf{C}^\dagger \mathbf{C}) \mathbf{S}_1 \mathbf{T}^\dagger \mathbf{S}_1^H \right. \\ &\quad \times (\mathbf{I}_J - \mathbf{C}^\dagger \mathbf{C}) \left. \right]^{-1} \\ &\quad \times \mathbf{S}_1 \mathbf{T}^\dagger (\mathbf{I}_{N-P} - \mathbf{S}_1^H \mathbf{C}^\dagger) \in \mathbb{C}^{J \times (N-P)}. \end{aligned} \quad (\text{C.14})$$

Expanding (C.12) yields

$$\mathbf{P}_S^\perp = \mathbf{P}_T^\perp (\mathbf{I} - \mathbf{S}_1^H (\mathbf{C}^\dagger + \mathbf{D})) = \mathbf{P}_T^\perp (\mathbf{I} - \mathbf{E}). \quad (\text{C.15})$$

From (5.24) and (5.25), the  $\hat{\mathbf{R}}_X$  can be rewritten as

$$\hat{\mathbf{R}}_X = \begin{bmatrix} \hat{\mathbf{R}}_{X,1} & \hat{\mathbf{R}}_{X,2} \\ \hat{\mathbf{R}}_{X,1}^H & \hat{\mathbf{R}}_{X,3} \end{bmatrix}, \quad (\text{C.16})$$

where

$$\hat{\mathbf{R}}_{X,1} = \hat{\mathbf{R}}_Y - \mathbf{Y}_0 \mathbf{P}_T^\perp \mathbf{E} \mathbf{Y}_0^H, \quad (\text{C.17})$$

$$\hat{\mathbf{R}}_{X,2} = \mathbf{Y}_0 \mathbf{P}_T^\perp (\mathbf{I} - \mathbf{E}) \mathbf{X}_{0,1}^H + \sum_{k=1}^K \mathbf{Y}_k \mathbf{X}_{k,1}^H, \quad (\text{C.18})$$

$$\hat{\mathbf{R}}_{X,3} = \mathbf{X}_{0,1} \mathbf{P}_T^\perp (\mathbf{I} - \mathbf{E}) \mathbf{X}_{0,1}^H + \sum_{k=1}^K \mathbf{X}_{k,1} \mathbf{X}_{k,1}^H. \quad (\text{C.19})$$

Applying the block matrix inversion lemma twice, first on  $\hat{\mathbf{R}}_X$  and then on  $\hat{\mathbf{R}}_{X,1}$ , we have

$$\hat{\mathbf{R}}_X^{-1} = \begin{bmatrix} \hat{\mathbf{R}}_Y^{-1} + \mathbf{W}_1 & \mathbf{W}_2 \\ \mathbf{W}_2^H & \mathbf{W}_3 \end{bmatrix} \quad (\text{C.20})$$

where

$$\mathbf{W}_1 = \hat{\mathbf{R}}_{X,1}^{-1} \hat{\mathbf{R}}_{X,2} \mathbf{W}_3 \hat{\mathbf{R}}_{X,2}^H \hat{\mathbf{R}}_{X,1}^{-1} \quad (\text{C.21})$$

$$\begin{aligned} & - \hat{\mathbf{R}}_Y^{-1} \mathbf{Y}_0 \left( \mathbf{I} + \mathbf{P}_T^\perp \mathbf{E} \mathbf{Y}_0^H \hat{\mathbf{R}}_Y^{-1} \mathbf{Y}_0 \right)^{-1} \\ & \times \mathbf{P}_T^\perp \mathbf{E} \mathbf{Y}_0^H \hat{\mathbf{R}}_Y^{-1}, \end{aligned}$$

$$\mathbf{W}_2 = - \hat{\mathbf{R}}_{X,1}^{-1} \hat{\mathbf{R}}_{X,2} \mathbf{W}_3, \quad (\text{C.22})$$

$$\mathbf{W}_3 = \left( \hat{\mathbf{R}}_{X,3} - \hat{\mathbf{R}}_{X,2}^H \hat{\mathbf{R}}_{X,1}^{-1} \hat{\mathbf{R}}_{X,2} \right)^{-1}. \quad (\text{C.23})$$

Inserting the above results in (5.28) followed by simple manipulations, we can see that the parametric GLRT test statistic (5.28) is equivalent to (5.29).

# Bibliography

- [1] J. Ward, "Space-time adaptive processing for airborne radar," Technical Report 1015, Lincoln Laboratory, MIT, December 1994.
- [2] R. Klemm, *Principles of Space-Time Adaptive Processing*, The Institute of Electrical Engineers, London, UK, 2002.
- [3] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 12–16, January 2002.
- [4] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, January 2002.
- [5] A. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, vol. 14, pp. 49–83, November 1997.
- [6] J. H. Justice, "Array processing in exploration seismology," in *Array Signal Processing*, S. Haykin, Ed. Prentice Hall, Englewood Cliffs, NJ, 1985.
- [7] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice Hall, Upper Saddle River, NJ, 1998.
- [8] L. E. Brennan and I. S. Reed, "Theory of adaptive radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 9, no. 2, pp. 237–252, 1973.
- [9] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 10, no. 6, pp. 853–863, 1974.
- [10] L. Cai and H. Wang, "On adaptive filtering with the CFAR feature and its performance sensitivity to non-Gaussian interference," in *Proceedings of the 24th Annual Conference on Information Sciences and Systems*, Princeton, NJ, March 1990, pp. 558–563.
- [11] W. Chen and I. S. Reed, "A new CFAR detection test for radar," *Digital Signal Processing*, vol. 1, no. 4, pp. 198–214, 1991.
- [12] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, January 1992.

- [13] E. J. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, pp. 115–127, March 1986.
- [14] L. T. McWhorter and L. L. Scharf, "Adaptive matched subspace detectors and adaptive coherence estimators," in *Proceedings of 30th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 1996, pp. 1114–1117.
- [15] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2538–2541, September 1999.
- [16] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 1–16, January 2001.
- [17] P. Li, H. Schuman, J. H. Michels, and B. Himed, "Space-time adaptive processing (STAP) with limited sample support," in *Proceedings of the 2004 IEEE Radar Conference*, April 2004, pp. 366–371.
- [18] M. Rangaswamy, B. Himed, and J. H. Michels, "Statistical analysis of the nonhomogeneity detector," in *Proceedings of the 2000 IEEE Radar Conference*, November 2000, vol. 2, pp. 1117–1121.
- [19] M. Rangaswamy, B. Himed, and J. H. Michels, "Performance analysis of the nonhomogeneity detector for STAP applications," in *Proceedings of the 2001 IEEE Radar Conference*, 2001, pp. 193–197.
- [20] M. Rangaswamy, J. H. Michels, and B. Himed, "Statistical analysis of the nonhomogeneity detector for non-gaussian interference backgrounds," in *Proceedings of the 2002 IEEE Radar Conference*, April 2002, pp. 304–310.
- [21] A. G. Jaffer, B. Himed, and P. T. Ho, "Estimation of range-dependent clutter covariance by configuration system parameter estimation," in *Proceedings of the 2005 IEEE Radar Conference*, May 2005, pp. 596–601.
- [22] M. Rangaswamy and J. H. Michels, "A parametric multichannel detection algorithm for correlated non-Gaussian random processes," in *Proceedings of the 1997 IEEE National Radar Conference*, Syracuse, NY, May 1997, pp. 349–354.
- [23] J. R. Román, M. Rangaswamy, D. W. Davis, Q. Zhang, B. Himed, and J. H. Michels, "Parametric adaptive matched filter for airborne radar applications," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 677–692, April 2000.
- [24] A. L. Swindlehurst and P. Stoica, "Maximum likelihood methods in radar array signal processing," *Proceedings of the IEEE*, vol. 86, pp. 421–441, February 1998.

- [25] J. Li, G. Liu, N. Jiang, and P. Stoica, "Moving target feature extraction for airborne high-range resolution phased-array radar," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 277–289, February 2001.
- [26] E. Conte and A. De Maio, "Distributed target detection in compound-Gaussian noise with Rao and Wald tests," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 2, pp. 568–582, April 2003.
- [27] A. De Maio, G. Alfano, and E. Conte, "Polarization diversity detection in compound-Gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 1, pp. 114–131, January 2004.
- [28] K. J. Sohn, H. Li, and B. Himed, "Parametric GLRT for multichannel adaptive signal detection," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5351–5360, November 2007.
- [29] S. L. Marple, Jr., *Digital Spectral Analysis with Applications*, Prentice Hall, Englewood Cliffs, NJ, 1987.
- [30] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, Englewood Cliffs, NJ, 1988.
- [31] T. Söderström and P. Stoica, *System Identification*, Prentice Hall International, London, UK, 1989.
- [32] R. Klemm, "Introduction to space-time adaptive processing," *Electronics and Communication engineering journal*, pp. 5–12, February 1999.
- [33] E. Conte, M. Lops, and G. Ricci, "Adaptive matched filter detection in spherically invariant noise," *IEEE Signal Processing Letter*, vol. 3, pp. 248–250, 1996.
- [34] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound Gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 2, pp. 617–625, April 1995.
- [35] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, Pearson Prentice Hall, Upper Saddle River, NJ, 2005.
- [36] T. Tsao, B. Himed, and J. H. Michels, "Effects of interference rank estimation on the detection performance of rank reduced stap algorithms," in *Proceedings of the 1998 IEEE Radar Conference*, May 1998, pp. 147–152.
- [37] H. Li, K. J. Sohn, and B. Himed, "The PAMF detector is a parametric Rao test," in *Proceedings of 39th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2005.

- [38] K. J. Sohn, H. Li, and B. Himed, "Multichannel parametric Rao detector," in *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2006)*, Toulouse, France, May 2006.
- [39] Y. Jiang, P. Stoica, and J. Li, "Array signal processing in the known waveform and steering vector case," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 23–35, January 2004.
- [40] K. J. Sohn, H. Li, and B. Himed, "Parametric Rao test for multichannel adaptive signal detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 920–933, July 2007.
- [41] J. Li, B. Halder, P. Stoica, and M. Viberg, "Computationally efficient angle estimation for signals with known waveforms," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2154–2163, September 1995.
- [42] D. Snegupta and S. M. Kay, "Parameter estimation and GLRT detection in colored non-Gaussian autoregressive processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1661–1676, October 1990.
- [43] K. J. Sohn, H. Li, B. Himed, and J. S. Markow, "Performance of multichannel parametric detectors with MCARM data," in *Proceedings of the 2007 IET International Conference on Radar Systems*, Edinburgh, UK, October 2007.
- [44] P. Wang, K. J. Sohn, H. Li, and B. Himed, "Performance evaluation of parametric Rao and GLRT detectors with KASSPER and bistatic data," in *Proceedings of the 2008 IEEE Radar Conference*, Rome, Italy, May 2008.
- [45] S. Lawrence Marple, Jr., P. M. Corbell, and M. Rangaswamy, "Multi-channel parametric estimator fast block matrix inverses," in *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2007)*, Honolulu, HI, April 2007, vol. 2, pp. 1137–1140.
- [46] K. J. Sohn, H. Li, and B. Himed, "Recursive parametric tests for multichannel adaptive signal detection," *IET Radar, Sonar and Navigation*, vol. 2, no. 1, pp. 63–70, February 2008.
- [47] S. Lawrence Marple, Jr., P. M. Corbell, and M. Rangaswamy, "Multi-channel fast parametric algorithms and performance for adaptive radar," in *Proceedings of 41st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2007, pp. 1835–1838.
- [48] S. Lawrence Marple, Jr., P. M. Corbell, and M. Rangaswamy, "Performance tradeoffs for multi-channel parametric adaptive radar algorithms," in *Proceedings of the 2008 International Conference on Radar*, September 2008, pp. 154–159.

- [49] Y. I. Abramovich, N. K. Spencer, and M.D.E. Turley, "Order estimation and discrimination between stationary and time-varying (TVAR) autoregressive models," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2861–2876, June 2007.
- [50] Y. I. Abramovich, N. K. Spencer, and M.D.E. Turley, "Time-varying autoregressive (TVAR) models for multiple radar observations," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1298–1311, April 2007.
- [51] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [52] J. S. Bergin and P. M. Techau, "High-fidelity site-specific radar simulation: KASSPER'02 workshop datacube," Technical Report ISL-SCRD-TR-02-105, Information Systems Laboratories, Inc., Vienna, VA, May 2002.
- [53] K. J. Sohn, H. Li, and B. Himed, "Parametric GLRT for multichannel adaptive signal detection," in *Proceedings of the 4th IEEE Workshop on Sensor Array and Multi-channel Processing (SAM06)*, Waltham, MA, July 12-14 2006.
- [54] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 281–295, February 1996.
- [55] K. J. Sohn, H. Li, and B. Himed, "Amplitude estimation of multichannel signal in spatially and temporally correlated noise," in *Proceedings of Progress In Electromagnetics Research Symposium 2006 (PIERS 2006)*, Cambridge, MA, March 2006.
- [56] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, Prentice Hall, Upper Saddle River, NJ, 1997.
- [57] Y. Zhang *et al.*, "Bistatic space-time adaptive processing (STAP) for airborne/spaceborne applications - Signal modeling and simulation tool for multichannel bistatic systems (SMS-MBS)," Technical Report AFRL-SN-RS-TR-2002-201, Part II (of II), Air Force Research Lab, Rome, NY, August 2002.
- [58] B. Himed, Y. Zhang, and A. Hajjari, "STAP with angle-doppler compensation for bistatic airborne radars," in *Proceedings of 2002 IEEE Radar Conference*, Long Beach, CA, April 2002, pp. 311–317.
- [59] B. Himed, "MCARM/STAP data analysis," Air Force Research Laboratory Final Technical Report, AFRL-SN-RS-TR-1999-48, Vol. II (of two), May 1999.
- [60] D. A. Landgrebe, *Signal Theory Methods in Multispectral Imaging*, Wiley, Hoboken, NJ, 2003.
- [61] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1392–1409, July 2001.



- [62] T. K. Sarkar, S. Park, J. Koh, and R. A. Schneible, "A deterministic least squares approach to adaptive antennas," *Digital Signal Processing*, , no. 6, pp. 185–194, 1996.
- [63] J. H. Michels, et al., "Multichannel parametric adaptive matched filter receiver," U.S. Patent No. 6,226,321, issued 1 May 2001.
- [64] J. H. Michels, et al., "Normalized parametric adaptive matched filter receiver," U.S. Patent No. 6,771,723, issued 3 August 2004.
- [65] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 19, pp. 1760–1770, October 1990.
- [66] D. Manolakis and D. Marden, "Non gaussian models for hyperspectral algorithm design and assessment," in *Proceedings of 2002 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'02)*, June 24–28 2002, vol. 3, pp. 1664–1666.
- [67] D. Marden and D. Manolakis, "Modeling hyperspectral imaging data," in *Proceedings of SPIE – Algorithms and Technologies for Multispectral, Hyperspectral and Ultraspectral Imagery*, 2003, vol. 5093, pp. 253–262.
- [68] D. Marden and D. Manolakis, "Using elliptically contoured distributions to model hyperspectral imaging data and generate statistically similar synthetic data," in *Proceedings of SPIE – Algorithms and Technologies for Multispectral, Hyperspectral and Ultraspectral Imagery X*, August 2003, vol. 5425, pp. 558–572.
- [69] A. D. Stocker and A. P. Schaum, "Application of stochastic mixing models to hyperspectral detection problems," in *Proceedings of SPIE – Algorithms for Multispectral and Hyperspectral Imagery III*, August 1997, vol. 3071, pp. 47–60.
- [70] G. Healey and D. Slater, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2706–2717, November 1999.
- [71] B. Thai and G. Healey, "Invariant subpixel material detection in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 3, pp. 599–608, March 2002.
- [72] J. H. Michels, B. Himed, and M. Rangaswamy, "Performance of STAP tests in Gaussian and compound-Gaussian clutter," *Digital Signal Processing*, vol. 10, no. 4, pp. 309–324, October 2000.
- [73] J. H. Michels, M. Rangaswamy, and B. Himed, "Performance of parametric and covariance based STAP tests in compound-Gaussian clutter," *Digital Signal Processing*, vol. 12, no. 2,3, pp. 307–328, April and July 2002.

- [74] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, CA, 1970.
- [75] H. Akaike, “A new look at statistical model identification,” *IEEE Transactions on Automatic Control*, vol. AC-19, pp. 716–723, 1974.
- [76] P. Chen, W. Melvin, and M. Wicks, “Screening among multivariate normal data,” *Journal of Multivariate Analysis*, vol. 69, pp. 10–29, 1999.
- [77] W. Melvin, “Space-time adaptive radar performance in heterogeneous clutter,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, pp. 621–633, April 2000.
- [78] J. H. Michels, B. Himed, and M. Rangaswamy, “Robust STAP detection in a dense signal airborne radar environment,” *Signal Processing*, special issue on New Trends and Findings in Antenna Array Processing for Radar, vol. 84, pp. 1625–1636, 2004.
- [79] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, NJ, 3 edition, 1996.
- [80] J. Y. Chen and I. S. Reed, “A detection algorithm for optical targets in clutter,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 23, no. 1, pp. 46–59, January 1987.
- [81] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.
- [82] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 3 edition, 1996.